



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 117384

TO: Manjunath N Rao
Location: REM/3B81/3C70
Art Unit: 1652
Monday, March 29, 2004

Case Serial Number: 09/847392

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Rao,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

117384

From: Rao, Manjunath N.
Sent: Monday, March 22, 2004 10:57 AM
To: STIC-Biotech/ChemLib
Subject: Sequence search request for 09/847,392

From: Manjunath N. Rao
Art Unit 1652, Room 3B81
Mail Box in Room 3C70
Phone: 272-0939

Date: 3-22-04

Please search the following as soon as possible for application with serial number

09/847,392

1. SEQ ID NO: 1 and nucleotides 557-1171 of SEQ ID NO:1 against all commercial nucleic acid databases, issued patents/published applications database and pending application database. Please provide a print of all results
2. SEQ ID NO: 2, against all commercial amino acid databases, issued patents/published applications database and pending application database. Please provide a print of all results.

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao
Art Unit 1652, Room 3B81,
Mail Box in Room 3C70,
Remsen Building, USPTO

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 3/24/04
Date Completed: 3/24/04
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH: 2
NA Sequences: _____
AA Sequences: 1 _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: 03/24/04
WWW/Internet: _____
Other (specify): _____

RECEIVED
MAR 22 2004
STIC

400⁺ Dulany St.
Alexandria, VA.
Phone: 571-272-0939

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case.



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg



GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:04:39 ; Search time 5020.17 Seconds
(without alignments)
10360.543 Million cell updates/sec

Title: US-09-847-392-1
Perfect score: 1200
Sequence: 1 agaaataatggagatgcg.....atgcggaatcggggtaaac 1200

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBank:
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_in.*
18: em_mu.*
19: em_mu.*
20: em_mu.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sta.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sv.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1200	100.0	1200	6	AR172781	AR172781 Sequence
2	1200	100.0	1200	6	E41573	E41573 DNA encodin
3	1200	100.0	1231	6	E48927	E48927 Novel gene
4	1200	100.0	1231	6	AX030175	AX030175 Sequence
5	1189	99.1	91414	1	ECOW85	M87049 E. coli gen
6	1180	98.3	11509	1	AE000458	AE000458 Escherich
7	1159.4	96.6	290976	1	AE016990	AE016990 Shigella
8	1146.6	95.5	313703	1	AE002567	AE002567 Escherich
9	1128.2	94.0	301566	1	AE016769	AE016769 Escherich
10	754.6	62.9	258050	1	AL627278	AL627278 Salmonell
11	754.6	62.9	293991	1	AE016845	AE016845 Salmonell
12	753	62.7	21692	1	AE008884	AE008884 Salmonell
13	753	62.7	96086	1	STYSTD1	AF233324 Salmonell
14	699	58.2	265383	2	AC020874	AC020874 Mus muscu
15	690	57.5	10253	1	AE015396	AE015396 Shigella
16	617	51.4	13301	1	AE005614	AE005614 Escherich
17	604.8	50.4	212936	2	AC020970	AC020970 Mus muscu
18	600.6	50.1	256373	2	AC020870	AC020870 Mus muscu
19	582.8	48.6	10592	1	AE005613	AE005613 Escherich
20	563	46.9	1576	1	ECPLDB	X03155 E. coli pld
21	527.8	44.0	9751	1	AE015397	AE015397 Shigella
22	417	34.8	417	6	AR352333	AR352333 Sequence
23	391.8	32.6	645	6	AR384166	AR384166 Sequence
24	361.4	30.1	10957	1	AE013640	AE013640 Yersinia
25	361.4	30.1	199050	1	AJ414159	AJ414159 Yersinia
26	297.2	24.8	1041	6	AR384122	AR384122 Sequence
27	294.4	24.5	214911	2	AC010537	AC010537 Homo sapi
28	270.4	22.5	268294	2	AC020885	AC020885 Mus muscu
29	243	20.2	214911	2	AC010537	AC010537 Homo sapi
30	233.2	19.4	208758	2	AC016129	AC016129 Drosophil
31	211.8	17.7	62274	2	AC020833	AC020833 Mus muscu
32	207.4	17.3	263	6	AR352166	AR352166 Sequence
33	191.4	16.0	63847	2	AC101459	AC101459 Mus muscu
34	165.4	13.8	11305	1	AE004104	AE004104 Vibrio ch
35	165.4	13.8	299938	1	AE016800	AE016800 Vibrio vu
36	162.6	13.5	245560	1	AF005330	AF005330 Vibrio vu
37	156	13.0	288108	1	AF005083	AF005083 Vibrio pa
38	146.8	12.2	3131	1	AB105408	AB105408 Photobact
39	144.8	12.1	1038	6	AR375232	AR375232 Sequence
40	138.4	11.5	110000	2	AC074221	Continuation (3 of
41	133	11.1	349907	1	BX571874	BX571874 Phototrab
42	112.4	9.4	10475	1	AE004937	AE004937 Pseudomon
43	107.6	9.0	313518	1	AE016856	AE016856 Pseudomon
44	91.4	7.6	11785	1	AE015905	AE015905 Shewanell
45	87.4	7.3	303226	1	AE016774	AE016774 Pseudomon

ALIGNMENTS

RESULT 1	AR172781	Sequence 1 from patent US 6303348.	DNA	linear	PAT 17-DEC-2001
LOCUS	AR172781	1200 bp			
DEFINITION	Sequence 1 from patent US 6303348.				
ACCESSION	AR172781				
VERSION	AR172781.1	GI:17912272			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 1200)				
AUTHORS	Livshits, V. Arkadievich., Zakataeva, N. Pavlovna., Aleoshin, V. Venyamovich., Balareova, A. Valentinovna. and Tokmakova, I. Lvovna				
TITLE	DNA coding for protein which confers on bacterium escherichia coli				

Pred. No. is the number of results predicted by chance to have a

resistance to L-homoserine and method for producing L-amino acids
Patent: US 6303348-A 1 16-OCT-2001;

JOURNAL
FEATURES
source
Location/Qualifiers
1. .1200
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 1200; DB 6; Length 1200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGGAGATCCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
DB 1 AGAAATAATGTGGAGATCCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
QY 61 GAGCCGGGTGAACTCTCTGCTGCCAGATCGCCAGATCATCAACATATCAATTAAG 120
DB 61 GAGCCGGGTGAACTCTCTGCTGCCAGATCGCCAGATCATCAACATATCAATTAAG 120
QY 121 CGATTAACTATGCCAGATCGCGATCGCTAAACAGCGACCGGAACGTCCTGCCCGCA 180
DB 121 CGATTAACTATGCCAGATCGCGATCGCTAAACAGCGACCGGAACGTCCTGCCCGCA 180
QY 181 TGGTCATGATTAAGACATCAACCCCAATGGAACAGGTATAGCCAGTTCGCGATAT 240
DB 181 TGGTCATGATTAAGACATCAACCCCAATGGAACAGGTATAGCCAGTTCGCGATAT 240
QY 241 TTTACGTAGCTCTCAATAGCCCGGCGAGATGACTACACCCGTCATGTGCTGTGCG 300
DB 241 TTTACGTAGCTCTCAATAGCCCGGCGAGATGACTACACCCGTCATGTGCTGTGCG 300
QY 301 CGAAACGACAAAGGCGACCGGATGTCATCCACAGTAACCTGTCTTATCAGC 360
DB 301 CGAAACGACAAAGGCGACCGGATGTCATCCACAGTAACCTGTCTTATCAGC 360
QY 361 TGACGCGAGAAATCACTAGCGGTCCTAGTAAAGCAGCAACGCTTTCTCTGTT 420
DB 361 TGACGCGAGAAATCACTAGCGGTCCTAGTAAAGCAGCAACGCTTTCTCTGTT 420
QY 421 TCCAGCTTTTGTGCTGAAACATCGGGTAATCTGCTCTTAAACACGATAAATCGT 480
DB 421 TCCAGCTTTTGTGCTGAAACATCGGGTAATCTGCTCTTAAACACGATAAATCGT 480
QY 481 TTTTGTAGCTGCTGACACACGCTGCGAGTAGAGTATGTGGCACAATAATAGAC 540
DB 481 TTTTGTAGCTGCTGACACACGCTGCGAGTAGAGTATGTGGCACAATAATAGAC 540
QY 541 ACACCGGAGTTTCATGACCTTGAATGTTGTTTGCCTACCTGTGACATCGATCAT 600
DB 541 ACACCGGAGTTTCATGACCTTGAATGTTGTTTGCCTACCTGTGACATCGATCAT 600
QY 601 TTTAAAGCTGTCGCCAGGCTCTGGTCAATCAACTATGACCACTCGCTCAACCCAGG 660
DB 601 TTTAAAGCTGTCGCCAGGCTCTGGTCAATCAACTATGACCACTCGCTCAACCCAGG 660
QY 661 TTATCCGCGCGTGGCGTCTATTTGCTGGCTTCAGACCGGATCGCATATGTC 720
DB 661 TTATCCGCGCGTGGCGTCTATTTGCTGGCTTCAGACCGGATCGCATATGTC 720
QY 721 TGGTGGCGTGGGTTGGGAGCTATTTCCGCTCAGTATGTTTGAAGTTGAA 780
DB 721 TGGTGGCGTGGGTTGGGAGCTATTTCCGCTCAGTATGTTTGAAGTTGAA 780
QY 781 GTGGGAGGCGCGGCTTACTTATTTGGCTGGGAATCCAGCAGTGGCGCGCTGTGTC 840
DB 781 GTGGGAGGCGCGGCTTACTTATTTGGCTGGGAATCCAGCAGTGGCGCGCTGTGTC 840
QY 841 AATTGACCTTAATCGTGGCTCTACTCAATCGCGTGCACATTTGTCAGCGCGAGT 900
DB 841 AATTGACCTTAATCGTGGCTCTACTCAATCGCGTGCACATTTGTCAGCGCGAGT 900
QY 901 TTTTGTGAATCTCAACCAATCCCAAAAGTATTTGTTTCTGGCGGCGCTATTTCCGCAAT 960

DB 901 TTTTGTGAATCTCAACCAATCCCAAAAGTATTTGTTTCTGGCGGCGCTATTTCCGCAAT 960
QY 961 CATCATGCCGCAACAGCGCACTGATGATGATATGCTGCTGGCGTCAACCTATTGT 1020
DB 961 CATCATGCCGCAACAGCGCACTGATGATGATATGCTGCTGGCGTCAACCTATTGT 1020
QY 1021 GTCATATATTATTGTGATGATGCTTACGCCACCTTGTCTCAACGGATTCCTATGGAT 1080
DB 1021 GTCATATATTATTGTGATGATGCTTACGCCACCTTGTCTCAACGGATTCCTATGGAT 1080
QY 1081 TAAAGCAACAGCAGATGAGGCGCTGAATAGATTTTCGGCTCGTTGTTATGCTGT 1140
DB 1081 TAAAGCAACAGCAGATGAGGCGCTGAATAGATTTTCGGCTCGTTGTTATGCTGT 1140
QY 1141 GGGAGCGCTGTTAGCATCGCGAGCGATGGCTGAAAATAATGTCGGATCGCGGTAAC 1200
DB 1141 GGGAGCGCTGTTAGCATCGCGAGCGATGGCTGAAAATAATGTCGGATCGCGGTAAC 1200

RESULT 2

E41573
LOCUS 1200 bp DNA linear PAT 31-JAN-2002
DEFINITION DNA encoding protein imparting tolerance against L-homoserine to Escherichia coli and process for producing L-amino acid.
ACCESSION E41573
VERSION E41573.1 GI:18627533
KEYWORDS JP 2000116390-A/1.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Rivishitsu, V.A.; Zakataeva, N.P.; Aryoshin, V.V.; Beraryo, A.V. and Tokumakova, I.R.
TITLE DNA encoding protein imparting tolerance against L-homoserine to Escherichia coli and process for producing L-amino acid
JOURNAL Patent: JP 2000116390-A 1 25-APR-2000;
COMMENT AJINOMOTO CO INC
OS Escherichia coli
PN JP 2000116390-A/1
PD 25-APR-2000
PF 12-OCT-1999 JP 1999289777
PR 13-OCT-1998 RU 98118425
PI VITARI ARUKAJEVICHI RIVISHITSU, NATARIYA PAVUROVUNA ZAKATAEVA,
PI VLADIMIR VENIYAMINOVICHI ARYOSHIN, ARA VALENTINOVUNA BERARYO,
PI IRINA RIVOVUNA TOKUMAKOVA
PC C12N15/09, C07K14/245, C12N1/21, C12P13/06, C12P13/08// (C12N1/21, C12R1:19),
PC C12N15/00
CC

PH Key Location/Qualifiers
FT CDS (557). (1171).
source
1. .1200
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"

ORIGIN

Query Match 100.0%; Score 1200; DB 6; Length 1200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
DB 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
QY 61 GAGCCGGGTGAACTCTCTGCTGCCAGATCGCCAGATCATCAACATATCAATTAAG 120
DB 61 GAGCCGGGTGAACTCTCTGCTGCCAGATCGCCAGATCATCAACATATCAATTAAG 120

121 CGATTAAATCCCGAGATCGGATCGCTAAACAGGCGACCGGAACGTCCTCCCGCGA 180
121 CGATTAAATCCCGAGATCGGATCGCTAAACAGGCGACCGGAACGTCCTCCCGCGA 180
181 TGGTCGATGATTAGACATCAACCCCAATGGACAGGTATAGCCAGTTCGGCATAT 240
181 TGGTCGATGATTAGACATCAACCCCAATGGACAGGTATAGCCAGTTCGGCATAT 240
241 TTACGTAGCTCTCAATACGCCCGGCGAGATGACTACCAACCGGTCATGTCCTGTGG 300
241 TTACGTAGCTCTCAATACGCCCGGCGAGATGACTACCAACCGGTCATGTCCTGTGG 300
301 CGAAACCGGACCAACCGGATGTCATCCACAGTAACTGTCGTCATCAACGC 360
301 CGAAACCGGACCAACCGGATGTCATCCACAGTAACTGTCGTCATCAACGC 360
361 TGACGCCGGAATCAGTCAGCGGTCCTATGTTAAAGCAGCAACGGTTTCTTTGTT 420
361 TGACGCCGGAATCAGTCAGCGGTCCTATGTTAAAGCAGCAACGGTTTCTTTGTT 420
421 TCCAGTCTTTTGTGCTGAAACATCGGTAATCTGCTCTTAAACCAAGTAAATCGT 480
421 TCCAGTCTTTTGTGCTGAAACATCGGTAATCTGCTCTTAAACCAAGTAAATCGT 480
481 TTTTGTAGCTGCTGACACCAACGCTGCGACAGTACGTAATGTCGACCAAAATAGAC 540
481 TTTTGTAGCTGCTGACACCAACGCTGCGACAGTACGTAATGTCGACCAAAATAGAC 540
541 ACACCGGAGTTCATGACCTTAGAATGTTGCTTACCTGACATCGATCAT 600
541 ACACCGGAGTTCATGACCTTAGAATGTTGCTTACCTGACATCGATCAT 600
601 TTTAAGCTGTGCGCAGCTCTGCTGAAATCAACACATGACACCTCGCTCAACACG 660
601 TTTAAGCTGTGCGCAGCTCTGCTGAAATCAACACATGACACCTCGCTCAACACG 660
661 TTATCCGCGGTCGCTATGCTGCGCTTACAGCCGACTGGCGATTCAATTGTGC 720
661 TTATCCGCGGTCGCTATGCTGCGCTTACAGCCGACTGGCGATTCAATTGTGC 720
721 TGGTTGGCGTGGGTTGGAGCTATTTCCGCTCAGTATGCTGTTGAAGTTGAA 780
721 TGGTTGGCGTGGGTTGGAGCTATTTCCGCTCAGTATGCTGTTGAAGTTGAA 780
781 GTGGCAGCGCGGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCTGCTGC 840
781 GTGGCAGCGCGGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCTGCTGC 840
841 AATTGACCTTAATCGCTGCGCTCTAATCGCTGCGATTTGTTCCAGCGCGCAGT 900
841 AATTGACCTTAATCGCTGCGCTCTAATCGCTGCGATTTGTTCCAGCGCGCAGT 900
901 TTTTGTGAATCTCACCAATCCAAAGTATGTTTCTGGCGGCTATTTCCGCAAT 960
901 TTTTGTGAATCTCACCAATCCAAAGTATGTTTCTGGCGGCTATTTCCGCAAT 960
961 CATATGCGGMAACAGCGCAATGATGAGTATATGCTGCGGCTACCACTATTGT 1020
961 CATATGCGGMAACAGCGCAATGATGAGTATATGCTGCGGCTACCACTATTGT 1020
1021 GGTGATATTTGATGATGCTGTTACGCCACCTTGTCTCAAGGATTCCTCTATGAT 1080
1021 GGTGATATTTGATGATGCTGTTACGCCACCTTGTCTCAAGGATTCCTCTATGAT 1080
1081 TAAAGGACCAAGCAGATGAAGGCGCTGAATTAAGATTTTCCGCTCGTTGTTATGCTGT 1140
1081 TAAAGGACCAAGCAGATGAAGGCGCTGAATTAAGATTTTCCGCTCGTTGTTATGCTGT 1140
1141 GGGAGCGTGTAGCATCGGCGAGCATGCGTGAAATTAATGTCGATCGCGCTAAAC 1200
1141 GGGAGCGTGTAGCATCGGCGAGCATGCGTGAAATTAATGTCGATCGCGCTAAAC 1200

RESULT 3
E48927
LOCUS 1231 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel gene and process for producing amino acid.
ACCESSION E48927
VERSION E48927.1 GI:18627996
KEYWORDS JP 2000189177-A/1.
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 1231)
AUTHORS Rivishitsu, V.A., Zakataeva, N.P., Aryoshin, V.V., Beraryo, A.V. and
Tokumakova, I.R.
TITLE Novel gene and process for producing amino acid
JOURNAL Patent: JP 2000189177-A 1 11-JUL-2000;
AJINOMOTO CO INC
COMMENT OS Escherichia coli
PN JP 2000189177-A/1
PD 11-JUL-2000
PF 15-DEC-1999 JP 1999356018
PR 23-DEC-1998 RU 98123511
PI VITARI ARUKAJEVIHI RIVISHITSU, NATARIYA PAVUROVUNA
ZAKATAEVA,
PI VLADIMIR VENIYAMINOVICHI ARYOSHIN, ARA VALENTINOVUNA BERARYO,
PI IRINA RIVOVUNA TOKUMAKOVA
PC C12N15/09, C07K14/245, C12N1/21, C12P13/06, C12P13/08,
PC C12P13/08//
PC C12N15/09, C12R1:19), (C12P13/06, C12R1:19), (C12P13/08, C12R1:19), PC
C12N15/00,
CC (C12N15/00, C12R1:19)
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..1231
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
FEATURES
source
ORIGIN
Query Match 100.0%; Score 1200; DB 6; Length 1231;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAATAATGTGGAGATCGCACCGCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 1 AGAATAATGTGGAGATCGCACCGCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
QY 61 GGACCGGCTGAACCTCTCTGCCAGATGCGCGCAGATCATCAACATAATCAATTAAG 120
Db 61 GGACCGGCTGAACCTCTCTGCCAGATGCGCGCAGATCATCAACATAATCAATTAAG 120
QY 121 CGATTACATGCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCTGCCCGCGA 180
Db 121 CGATTACATGCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCTGCCCGCGA 180
QY 181 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAAACAGGTATAGCCAGTTCGGCATAT 240
Db 181 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAAACAGGTATAGCCAGTTCGGCATAT 240
QY 241 TTTACGTAGCTCTCAATACGCCCGGCGAGATGACTACCAACCGGTCATGTCCTGTGG 300
Db 241 TTTACGTAGCTCTCAATACGCCCGGCGAGATGACTACCAACCGGTCATGTCCTGTGG 300
QY 301 CGAAACCGGACCAACCGGATGTCATCCACAGTAACTGTCGTCATCAACGC 360
Db 301 CGAAACCGGACCAACCGGATGTCATCCACAGTAACTGTCGTCATCAACGC 360
QY 361 TGACGCCGGAATCAGTCAGCGGTCCTATGTTAAAGCAGCAACGGTTTCTTTGTT 420
Db 361 TGACGCCGGAATCAGTCAGCGGTCCTATGTTAAAGCAGCAACGGTTTCTTTGTT 420

QY 421 TCCAGCTCTTTTGTCTGTAACATCGGTAATCTGCTCTTAAACACACGTAATAATCGT 480
DB 421 TCCAGCTCTTTTGTCTGTAACATCGGTAATCTGCTCTTAAACACACGTAATAATCGT 480
QY 481 TTTTATTAGCTGCTGACACAAACGCTGCGAGTAGGTAATGTGGCACAATAATAGAC 540
DB 481 TTTTATTAGCTGCTGACACAAACGCTGCGAGTAGGTAATGTGGCACAATAATAGAC 540
QY 541 ACACCGGAGTTTCATCATCACTTGAATGGTGGTTTGCCTTACCTGCTGACATGATCAT 600
DB 541 ACACCGGAGTTTCATCATCACTTGAATGGTGGTTTGCCTTACCTGCTGACATGATCAT 600
QY 601 TTTAACGCTGTGCCAGGCTGTGGTGCATCAACTAATCACTATGACCACTGCTCAACACCG 660
DB 601 TTTAACGCTGTGCCAGGCTGTGGTGCATCAACTAATCACTATGACCACTGCTCAACACCG 660
QY 661 TTATCCGCGCGGTGGCTCTATTGTGGGCTTCAGACCGGACTGGCATATCATTTGTC 720
DB 661 TTATCCGCGCGGTGGCTCTATTGTGGGCTTCAGACCGGACTGGCATATCATTTGTC 720
QY 721 TGGTTGGCGTGGGTTGGAGCCTATTTTCCGCTCACTGATTCGTTTGAAGTGTGAA 780
DB 721 TGGTTGGCGTGGGTTGGAGCCTATTTTCCGCTCACTGATTCGTTTGAAGTGTGAA 780
QY 781 GTGGCAGCGCGGCTTACTTGAATGGCTGGATCCAGACAGTGGCGCGCTGCTGTC 840
DB 781 GTGGCAGCGCGGCTTACTTGAATGGCTGGATCCAGACAGTGGCGCGCTGCTGTC 840
QY 841 AATTGACCTTAAATCGCTGGCTCTACTCAATCGGCTGCAATTTGTTTCAGCGCGCAGT 900
DB 841 AATTGACCTTAAATCGCTGGCTCTACTCAATCGGCTGCAATTTGTTTCAGCGCGCAGT 900
QY 901 TTTTGTGAATCTCAACATCCCAAAAGTATGTGTTTCTGGCGGCTATTTCCGGAATT 960
DB 901 TTTTGTGAATCTCAACATCCCAAAAGTATGTGTTTCTGGCGGCTATTTCCGGAATT 960
QY 961 CATCATGCGCAACAGCGCACTGATGAGTATATCGTCTGCGGCTGACCACTATTGT 1020
DB 961 CATCATGCGCAACAGCGCACTGATGAGTATATCGTCTGCGGCTGACCACTATTGT 1020
QY 1021 GGTGATATTATGTGATGATCGGTAGCCACCTTCTGCTCAACGGAATGCTCTATGAT 1080
DB 1021 GGTGATATTATGTGATGATCGGTAGCCACCTTCTGCTCAACGGAATGCTCTATGAT 1080
QY 1081 TAAAGACCAAGCAGATGAAGCGCTGAATGAATTTTGGCTGCTTTTATGCTGCT 1140
DB 1081 TAAAGACCAAGCAGATGAAGCGCTGAATGAATTTTGGCTGCTTTTATGCTGCT 1140
QY 1141 GGGAGCGCTTTAGCATCGCGAGGATCGCTGAAAAATAATGTGCGATGCGCGTAAAC 1200
DB 1141 GGGAGCGCTTTAGCATCGCGAGGATCGCTGAAAAATAATGTGCGATGCGCGTAAAC 1200

RESULT 4
AX030175 LOCUS 1231 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent EP1013765.
ACCESSION AX030175
VERSION AX030175.1 GI:10190392
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1. Belareva, A.V., Aleshin, V.V., Livshits, V.A., Tokhmakova, I.L. and Zakataeva, N.P.
AUTHORS Gene and method for producing l-amino acids
TITLE Patent: EP 1013765-A 1 28-JUN-2000;
JOURNAL ATINOMOTO KK (JP)
FEATURES Location/Qualifiers
source 1..1231

/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
557..1174
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09318.1"
/db_xref="GI:10190393"
/translation="MTLEWFFAYLLLSILTLSPGSGAINTMTTSLNHGVPAGGVYCM
ASDTGSIYAGMRGVTFLFSVIAPFVLLKWAAGAYLLIWLGIQNRAGAIDKLSLA
STQRRLLFRFAVNLNPKSIVFLAAPPFQIPQPOLMQYIVLGVTTIVVDIIV
MIGVATLAQRALWIKGPKQMKALKIKFSGLEFLVGLLASARHA"

ORIGIN

Query Match 100.0%; Score 1200; DB 6; Length 1231;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAATAATGTGGAGATGCGACCCGCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
DB 1 AGAATAATGTGGAGATGCGACCCGCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
QY 61 GGACCGGCTCAACTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 120
DB 61 GGACCGGCTCAACTCTCTGCTGCCAGATGCCGCCAGATCATCAACATAATCATTAAG 120
QY 121 CGATTAAATCCCGAGATGCGGCTAAGCAGCGACCGGAAGCTCCCTCCCGCGGA 180
DB 121 CGATTAAATCCCGAGATGCGGCTAAGCAGCGACCGGAAGCTCCCTCCCGCGGA 180
QY 181 TGGTCGATGATTAGACATCAACCCCAATGGAACAGGTCTAGGCCAGTTCGGCATAT 240
DB 181 TGGTCGATGATTAGACATCAACCCCAATGGAACAGGTCTAGGCCAGTTCGGCATAT 240
QY 241 TTTAGTAGCTCTCAATACGCGCGGCGAGATGACTACCAACGCGGTCTGCTGTGG 300
DB 241 TTTAGTAGCTCTCAATACGCGCGGCGAGATGACTACCAACGCGGTCTGCTGTGG 300
QY 301 CGAAAAACGCAAGCGGACCGGAATGTCTACACACAGTAACCTCTGCTTCAATCAGC 360
DB 301 CGAAAAACGCAAGCGGACCGGAATGTCTACACACAGTAACCTCTGCTTCAATCAGC 360
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCTCATGTAAGCAGCAACGCGTTTCTCTTGT 420
DB 361 TGACGCCAGAAATCAGTCAGCGGTCCTCATGTAAGCAGCAACGCGTTTCTCTTGT 420
QY 421 TCCAGTCTTTTGTGCTGAAACATCGGTAATCTGCTCTTTAAACCGACGTAATTCGT 480
DB 421 TCCAGTCTTTTGTGCTGAAACATCGGTAATCTGCTCTTTAAACCGACGTAATTCGT 480
QY 481 TTTTATTAGCTGCTGACACACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 540
DB 481 TTTTATTAGCTGCTGACACACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 540
QY 541 ACACCGGAGTTTCATCATCACTTGAATGGTGGTTTGCCTTACCTGCTGACATGATCAT 600
DB 541 ACACCGGAGTTTCATCATCACTTGAATGGTGGTTTGCCTTACCTGCTGACATGATCAT 600
QY 601 TTTAACGCTGTGCCAGGCTGTGGTGCATCAACTAATCACTATGACCACTGCTCAACACCG 660
DB 601 TTTAACGCTGTGCCAGGCTGTGGTGCATCAACTAATCACTATGACCACTGCTCAACACCG 660
QY 661 TTATCCGCGCGGTGGCTCTATTGTGGGCTTCAGACCGGACTGGCATATCATTTGTC 720
DB 661 TTATCCGCGCGGTGGCTCTATTGTGGGCTTCAGACCGGACTGGCATATCATTTGTC 720
QY 721 TGGTTGGCGTGGGTTGGAGCCTATTTTCCGCTCACTGATTCGTTTGAAGTGTGAA 780
DB 721 TGGTTGGCGTGGGTTGGAGCCTATTTTCCGCTCACTGATTCGTTTGAAGTGTGAA 780
QY 781 GTGGCAGCGCGGCTTACTTGAATGGCTGGATCCAGACAGTGGCGCGCTGCTGTC 840


```

VNLFRHNRNRLTAAGKLLPYAETLMTWQARKWRIPHDITSLFSLVPAPRCGNV
CLISWDACIKIRWPIQAYSSKRLPNGLM"
Complement (1136. .1165)
/notes="promoter matrix score of 63"
1151. .1179
/notes="promoter matrix score of 57"
1223. .1636
/genes="ol37"
1223. .1636
/genes="ol37"
/genes="ol37"
/function="unknown"
/codon_start=1
/transl_table=11
/protein_id="AAA67569.1"
/db_xref="GI:148172"
/translation="WAESFTTNRYFDNKHVPRGSRHGDTTIKEAQLLHGHYATNE
LDLGRREVTEEEKLFVAVCGEREPEVTEAEKWSKYMTIRIKPKRFHTLSAVNRRLK
VLKTPITLTKNGEMPPFYVSSPCRWMSKRSIAR"
1562. .1586
/genes="ol37"
/notes="terminator-like sequence"
complement (1585. .3135)
/genes="f516"
complement (1585. .3135)
/genes="f516"
/function="unknown"
/notes="similar to Mg chelatase subunit of Rhodobacter
capsulata"
/codon_start=1
/transl_table=11
/protein_id="AAA67569.1"
/db_xref="GI:148173"
/translation="WHIVTGGGLMSLSIVTRAALGVNAPPIVTHVHSKGLPLTLM
VGLPETTKKARDVRISAIINSYEPKAKITINLAPDLPKGGRYDLDPIALALAA
SQLTANKLDEYELGEALTGALRGVPGATISSATEAIKSGRKIIIVAKDNEDVGLIN
GGECILIAHLOQVACAFLEKALJEXPKPTDAVSRALOHLDSDVIGQGRKGLIETAA
GSHNLLIGPPTGKTMLASRINGLLPDSNEALSAAILSLVNAESVOKWRQPPF
RSPHSASLTAMVGGGATPGRGESISLAHNGVLFDELPEPERRTLDALEPIESGOIH
LSRTPAKITYPARFOLVAMNPSPTGHVQGNHNECTPEQTLRVNLRLSGPFLDRFLS
LSIPAPFGLISKTVPGESATVKQVRMAARQPKRQKNLNLWLDSPISROFCKLE
SEDAMWLEGLIHLGLSRAWORLLKVAITDIQSDIITROHLEQAVSYRAIDRL
IHLQKLLT"
3287. .3433
/notes="predicted bend of 79 degrees"
3297. .3326
/notes="promoter matrix score of 74"
3336. .12791
/notes="corresponds to M10313 (1. .9456)"
3336. .3737
/notes="corresponds to V00289 (1. .402)"
3336. .3676
/notes="corresponds to M11651 (1. .341)"
3379. .5721
/notes="V00290; ECILVX (1. .2343)"
3458. .3556
/genes="ilvL"
/genes="ilvL"
/codon_start=1
/transl_table=11
/product="ilvGMDA operon leader peptide"
/protein_id="AAA67570.1"
/db_xref="GI:148174"
/translation="MTALLRVISLVVISVWVIIIPPCGAALRGKA"
3696. .4679
/genes="ilvG"
3696. .4679
/genes="ilvG"
/notes="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=11
/product="acetohydroxy acid synthase II, large subunit"
/protein_id="AAA67571.1"

```

```

/db_xref="GI:148175"
/translation="MNGAOWVHVALRAOGVNTVFGVPGGAIIMPVVDALYDGGVEHLLC
RHEGGAARAIQYARATGTCVCIATLITGLADALLDSIPVVAITGOVSAP
FIGTDAFOEVDVLGLSLACTKSFIVQSEELPRINAEAPDVACSRPGVLDVLPKD
IQAASGLEPFWTIVENEVAFPHAEVQARQMLAKAKPMNVYGGVGMVAQVPAIRE
FLAATKMPATCTILKLGAVEADYFYLGMGHGTGAANFAVQECDLLLTAQAFDRR
VTGKLANTFAPHASVVIHMDIDPAEMNKLRQAHVALQGLDNALLPLAQOQLNQ"
4675. .5340
/genes="o221"
4675. .5340
/genes="o221"

Query Match 99.1%; Score 1189; DB 1; Length 91414;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
DB 62717 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 62658

QY 61 GGACCGGCTGAACCTCTCTGTCGCCAGATGCGGCCAGATCATCAACATAATCATTAAG 120
DB 62657 GGACCGGCTGAACCTCTCTGTCGCCAGATGCGGCCAGATCATCAACATAATCATTAAG 62598

QY 121 CGATTACATGCCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCCTCCCGCGCA 180
DB 62597 CGATTACATGCCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCCTCCCGCGCA 62538

QY 181 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTTCATAGCCAGTTCGCCATAT 240
DB 62537 TGGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTTCATAGCCAGTTCGCCATAT 62478

QY 241 TTTAGTAGTCTCTCAATACGCCCGGCGAGATGACTACACCGGTCATGGTCTGTGG 300
DB 62477 TTTAGTAGTCTCTCAATACGCCCGGCGAGATGACTACACCGGTCATGGTCTGTGG 62418

QY 301 CGAAAAACGGAACAAAGCCGCAAGTCTATCCACACAGTAACTCTGCTTCATCAGC 360
DB 62417 CGAAAAACGGAACAAAGCCGCAAGTCTATCCACACAGTAACTCTGCTTCATCAGC 62358

QY 361 TGACCCAGAAATCAGTCAGGCTCCATGTTAAAGCAGCAACCGGTTCTCTGTT 420
DB 62357 TGACCCAGAAATCAGTCAGGCTCCATGTTAAAGCAGCAACCGGTTCTCTGTT 62298

QY 421 TCCAGTCTTTTGTCTGTAACATCGGTATCTGCTCTTTAAACACAGTAAATCGT 480
DB 62297 TCCAGTCTTTTGTCTGTAACATCGGTATCTGCTCTTTAAACACAGTAAATCGT 62238

QY 481 TTTTTHAGCTGCTGACACACACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 540
DB 62237 TTTTTHAGCTGCTGACACACACGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 62178

QY 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGTTCCTACTCTGTCGACATCGATCAT 600
DB 62177 ACACCGGAGTTTCATCATGACCTTAGAATGGTGTTCCTACTCTGTCGACATCGATCAT 62118

QY 601 TTTAACGCTGTGCCAGGCTCTGCTGCAATCAACATATGACCACTCGCTCGCTCAACACGG 660
DB 62117 TTTAACGCTGTGCCAGGCTCTGCTGCAATCAACATATGACCACTCGCTCGCTCAACACGG 62058

QY 661 TTATC-CGGCCGGTGGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATTTGTG 719
DB 62057 TTATC-CGGCCGGTGGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATTTGTG 61998

QY 720 CTGCTGGCGTGGGCTTGGGACGCTATTTTCCGCTCAGTGATGGCTTTGAAGTGTGA 779
DB 61997 CTGCTGGCGTGGGCTTGGGACGCTATTTTCCGCTCAGTGATGGCTTTGAAGTGTGA 61938

QY 780 AGTGGGAGGCGCGCTTACTTGTATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTG 839
DB 61937 AGTGGGAGGCGCGCTTACTTGTATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTG 61878

QY 840 CAATTGACCTTAAATCGTGGCCTCTACTCAATCGCTCGACATTTTGTTCAGCGCGCAG 899

```

```
Db 61877 CAATTGACCTTAATCGTGGCTTACTCAATCGCTGCACATTGTTTCAGCGCGCAG 61818
QY 900 TTTTGTGAATCTACCAATCCCAAAAGATATGTGTTTCTGGCGGCGCTATTTCCGCAAT 959
Db 61817 TTTTGTGAATCTACCAATCCCAAAAGATATGTGTTTCTGGCGGCGCTATTTCCGCAAT 61758
QY 960 TCATCATGCGCAACAGCGGCACTGATGCGATATGCTGCTCGCGTCAACCACTATTG 1019
Db 61757 TCATCATGCGCAACAGCGGCACTGATGCGATATGCTGCTCGCGTCAACCACTATTG 61698
QY 1020 TGGTCATATTATTTGATGATCGTTTACCGCACCTTGTCTCAACGGATTGCTCTATGGA 1079
Db 61697 TGGTCATATTATTTGATGATCGTTTACCGCACCTTGTCTCAACGGATTGCTCTATGGA 61638
QY 1080 TTAAGAGCCAAAGCAGATGAGGCGCTGAATAGATTTTGGCTCGCTGTTATGCTGG 1139
Db 61637 TTAAGAGCCAAAGCAGATGAGGCGCTGAATAGATTTTGGCTCGCTGTTATGCTGG 61578
QY 1140 TGGAGCGCTGTTAGCATCGGCGAGGCGATGCGTGAAAAATTAATCTCGATGCGCGCTAAA 1199
Db 61577 TGGAGCGCTGTTAGCATCGGCGAGGCGATGCGTGAAAAATTAATCTCGATGCGCGCTAAA 61518
QY 1200 C 1200
Db 61517 C 61517

RESULT 6
LOCUS AE000458/c
DEFINITION Escherichia coli K12 MG1655 section 348 of 400 of the complete genome.
VERSION AE000458 U00096
KEYWORDS 97426617
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
REFERENCE 1 (bases 1 to 11509)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 11509)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 3 (bases 1 to 11509)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
REFERENCE 4 (bases 1 to 11509)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT On Sep 9, 1997 this sequence version replaced gi:1790254. This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently
```

determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amher.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes are periodically updated; updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
sourceLocation/Qualifiers
1. 11509

/organism="Escherichia coli K12"

/mol_type="genomic DNA"

/strain="K12"

/sub_strain="MG1655"

/db_xref="taxon:83333"

81. 950

/gene="pldA"

/note="synonym: b3821"

81. 950

/gene="pldA"

/EC_number="3.1.1.32"

/function="enzyme; Central intermediary metabolism:

Phosphorus compounds"

/note="o289; 99 pct identical to PAI_ECOLI SW: P00631;

CGSCNO. 384"

/codon_start=1

/transl_table=11

/product="outer membrane phospholipase A"

/protein_id="AAC76824.1"

/db_xref="GI:2367300"

/translation="MRTLGQLLPVFMPLMVAQATVKEVHDPAVGGSIANNMLQ
EHNPFLLPYDNTNLYITQSDLNKEAIAASYDMAENARKDEVKFQSLAFLMVGIL
GPNSVLGASYTQKSWQLSNSESSPFRNTEYQPLFLGFDATYRPAQWTLRDVEMGY
NHDNRSRSDPTSRNRLYTRLMANGNMLVEKVPWYVGVNTDDNDPDIKYMGYYOLK
IGYHGLDAVLGSAKQYNWNTGYGAGLGLSLYPITKHVRLYTVYSGYSGSLIDYNENQ
TRVGQVGLNDLP"

965. 993

/note="factor Sigma70; predicted +1 start at 4003392"

1019. 1047

/note="factor Sigma70; predicted +1 start at 4003446"

1077. 2909

/gene="recQ"

/note="synonym: b3822"

1077. 2909

/gene="recQ"

/EC_number="3.6.1.-"

/function="enzyme; DNA - replication, repair,

restriction/modification"

/note="o610; 99 pct identical to 607 amino acids

of RECQ_ECOLI SW: P15043 but has 3 additional N-ter

residues"

/codon_start=1

/transl_table=11

/product="ATP-dependent DNA helicase"

/protein_id="AAC76825.1"

/db_xref="GI:2367301"

/translation="MNVQAQAEVLNLSGAKQVLOETFGYQQRPGQERTIDTVLSGRD
CLVMPFGGKSLCYQIPALLNLGLTVVVSPLSLMKQDQQLQANGVAACLNSQTQ

Db	3989	GCATTGACCTTAATCGCTGGCTCTACTCAATCGCTGCACATTGTTCCAGCGCCA	3930
QY	899	GTTTTTGGAATCTCAACAATCCAAAAGTATTGTTTCTGGCGCGCTATTTCCGCAA	958
Db	3929	GTTTTTGGAATCTCACCAATCCAAAAGTATTGTTTCTGGCGCGCTATTTCCGCAA	3870
QY	959	TTTCATCTGCGCGCAACAGCGCAACTGATGCGAGTATATCGTCTCGGCGTCACCACTATT	1018
Db	3869	TTTCATCTGCGCGCAACAGCGCAACTGATGCGAGTATATCGTCTCGGCGTCACCACTATT	3810
QY	1019	GTGTGTCGATATATTGTGATGATCGGTTACGCCACCTTGCTCAACGGANTGCTCTAAG	1078
Db	3809	GTGTGTCGATATATTGTGATGATCGGTTACGCCACCTTGCTCAACGGANTGCTCTAAG	3750
QY	1079	ATTAAGGACCAACAGCAGATGAAGGCGCTGAATAAGATTTTCGGCTCGTGTGTTATGCTG	1138
Db	3749	ATTAAGGACCAACAGCAGATGAAGGCGCTGAATAAGATTTTCGGCTCGTGTGTTATGCTG	3690
QY	1139	GTGGGAGCGCTGTTAGCATCGGCGAGCGATCGTGAAAAATATTCGGATCGCGCGTAA	1198
Db	3689	GTGGGAGCGCTGTTAGCATCGGCGAGCGATCGTGAAAAATATTCGGATCGCGCGTAA	3630
QY	1199	AC	1200
Db	3629	AC	3628
RESULT 7			
LOCUS	AE016990		
DEFINITION	Shigella flexneri 2a str. 290976 bp DNA linear BCT 22-APR-2003		
ACCESSION	AE016990	AE014073	
VERSION	AE016990.1	GI:30042943	
KEYWORDS	Shigella flexneri 2a str. 2457T		
SOURCE	Shigella flexneri 2a str. 2457T		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.		
REFERENCE	1 (bases 1 to 290976)		
AUTHORS	Wei J., Goldberg, M.B., Burland V., Venkatesan, M.M., Deng, W., Fournier, G., Maynew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.		
TITLE	Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T		
JOURNAL	Infect. Immun. 71 (5), 2775-2786 (2003)		
PUBMED	12704152		
REFERENCE	2 (bases 1 to 290976)		
AUTHORS	Wei J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Maynew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA		
FEATURES	Location/Qualifiers		
source	1. 290976		
	/organism="Shigella flexneri 2a str. 2457T"		
	/mol_type="genomic DNA"		
	/strain="2457T"		
	/serotype="2a"		
	/db_xref="taxon:198215"		
gene	complement(68..1405)		
	/gene="yJdE"		
	/locus_tag="S3622"		
CDS	complement(68..1405)		
	/gene="yJdE"		
	/locus_tag="S3622"		
	/function="putative transport; Not classified"		
	/note="residues 1 to 445 of 445 are 92.35 pct identical to residues 1 to 445 of 445 from Escherichia coli K-12 ; residues 1 to 445 of 445 from Escherichia coli K-12 ;		

```

B4115"
/codon_start=1
/transl_table=1
/product="putative amino acid/amine transport protein"
/protein_id="AAP18666.1"
/db_xref="GI:3004294"
/translation="MSSDADAHKVGKLVPTLTVMSNGINSGVFLFLPANLASTGIAIY
GMLVTIICALGLSVMYAKMSFLDPSPGSYAYARCFGFLGYQTNVLYMLACWIGNI
AMVILGVGLYVFPILKDLVLITITCVVLI FVLLNIVGPKMITRQVATVFLVI
PIVGIYAGPWFPEGTGMAVNNVSLGFGALQSTLWLSFIGVESASVAAGVVK
NPKRNPVATIGVLLIYAAVCYVLSTTAINGMIPNAALVSASPFGDAARVALGDTAGA
IVSCRAAGCGUGGWTLAGGTAKAADADGFLPPFARVNKAGTVPAGLIIIVGLIM
TIFQLSISPNAKEFGVESVITFLVFLYITCAALLLIGHGHGFKARAPLAKSD"
/complement(1542..2303)
/gene="adiY"
/locus_tag="S3623"
/complement(1542..2303)
/gene="adiY"
/locus_tag="S3623"
/complement(1542..2303)
/function="putative regulator; Not classified"
/notes="residues 1 to 253 of 253 are 99.28 pct identical to
residues 1 to 253 of 253 from Escherichia coli K-12 :
B4116"
/codon_start=1
/transl_table=1
/product="putative ARAC-type regulatory protein"
/protein_id="AAP18667.1"
/db_xref="GI:30042945"
/translation="MRCSDQPCIVLLTPEKQVIRVNGKEPISUKANHMALLNCENNI
IVSSLLNLTVAHISHDIIKDYLRFNKDLQIPVQRSATPILTPCLTPDFVFLVAA
QISMNPAETSEKERTALLFTVLSFLDKKFLSMYMLRNCSVSVQIIIESDIH
KGNLISVATSCILSPCLKKLLKSLKNTSISOIITTCRMVAVANLMMDKGKISQVSQ
SOWNVSTSYFISVFKDYGWTPILHYYSQHRRTVA"
/complement(2628..4895)
/gene="adiA"
/locus_tag="S3624"
/complement(2628..4895)
/gene="adiA"
/locus_tag="S3624"
/function="enzyme; Degradation of small molecules: Amino
acids."
/notes="residues 1 to 755 of 755 are 99.47 pct identical to
residues 2 to 756 of 756 from Escherichia coli K-12 :
B4117"
/codon_start=1
/transl_table=1
/product="biodegradative arginine decarboxylase"
/protein_id="AAP18668.1"
/db_xref="GI:30042946"
/translation="MKVLI VESFLHDTWGVNAVERLADALSOQNVTVIKTSFDDG
FALLSNBAICIMFSQMBPEHONRQLLKLHERQNVYFLIGDSEKALAWIGNI
ROLLELDEFWILLEDITADIAGRAVAMTRIQQLLPFLFSALMKXSDIHEYSWAP
GHQGVGFTKTPAGFYVHDYIGENLFTDVRNCHKSIEQGLMLTGAKPVCNAKQ
GADQSVNVTGSGNSTINMQACMTDNVDDVDRNCHKSIEQGLMLTGAKPVCNAKQ
NRYGIIGIYFQENQPTLQKTI SESLFLPKDAGKPGSYCVTVNTGYDGVICYNKQ
DLLEKTSORLHFAGYAFNPVIYADHYAMRGEPGDHNGPTVFATHSIHLNLALS
QASYIYHVEGREGAFNFRQNA YMHATSPLYIACSNDAVSMMDGNSGLITQEV
IDEADYDQMARLYKEFTADSGSFPPKNEKVTDPTQKTYDFADAPTKLITQVD
CWMHPGSGWGFKDIDFNWMSLDPIKSIILAPGMGEDGELEETGVPAALVATLWLRG
VIYPTRTIDFQIMFLFMSWGKGLNVLNLSFKHIDYANTFLAQMVELVEQDFP
GIATVNTIDGLDGTWFAWKENNPAGNLNAYSGLPVABITPREAYNAIVONNVELVSI
ENLPGRIAANSVIPYFGGIPMLLSGNGFNGDKNSQVSYRLSLQSDWHHFFRFPEHE"
/complement(5094..6002)
/gene="melR"
/locus_tag="S3625"
/complement(5094..6002)
/gene="melR"
/locus_tag="S3625"
/function="regulator; Degradation of small molecules:
Carbon compounds"
/notes="residues 1 to 302 of 302 are 99.66 pct identical to

```


Db 252154 CGCAATTGATCTTAATCGCTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGC 252213
QY 898 AGTTTGTGGAATCTCACCATCCCAAGATTTGTTTCTGCGGCGCTATTTCCGCA 957
Db 252214 AGTTTGTGGAATCTCACCATCCCAAGATTTGTTTCTGCGGCGCTATTTCCGCA 252273
QY 958 ATTATCATCTCGCGCAACGCGCAACTGATCGATATATCGTCTCGCGCTCACCACAT 1017
Db 252274 ATTATCATCTCGCGCAACGCGCAACTGATCGATATATCGTCTCGCGCTCACCACAT 252333
QY 1018 TGTGTCGATATATGATGATCGGTAGCCACCCCTTCTCAACGGATTCCTCTATG 1077
Db 252334 TGTGTCGATATATGATGATCGGTAGCCACCCCTTCTCAACGGATTCCTCTATG 252393
QY 1078 GATTAAAGACCAACGAGATGAAGCGCTCAATAAGATTTTCGGCTCGTTGTTTATGCT 1137
Db 252394 GATTAAAGACCAACGAGATGAAGCGCTCAATAAGATTTTCGGCTCGTTGTTTATGCT 252453
QY 1138 GTGGGAGCGCTGTAGCATCGCGAGGCGATCGTGAAATAATATGTCGATCGCGCGTA 1197
Db 252454 GTGGGAGCGCTGTAGCATCGCGAGGCGATCGTGAAATAATATGTCGATCGCGCGTA 252513
QY 1198 AAC 1200
Db 252514 AAC 252516

RESULT 8
AP002567/c
LOCUS 318703 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 18/20.
ACCESSION AP002567 BR000007
VERSION AP002567.1 GI:13364198
KEYWORDS
SOURCE Escherichia coli O157:H7
ORGANISM Escherichia coli O157:H7
REFERENCE
AUTHORS Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
10734605
REFERENCE
AUTHORS Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
11108008
REFERENCE
AUTHORS Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and Shinagawa, H.
Complete nucleotide sequence of the prophage VT4-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
11111050
REFERENCE
AUTHORS Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,

Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
11258796
REFERENCE
AUTHORS 5 (bases 1 to 318703)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp)
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)
genome project.
COMMENT
FEATURES
Location/Qualifiers
1. 318703
/organism="Escherichia coli O157:H7"
/mol_type="genomic DNA"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
103..648
/gene="ECs4723"
103..648
/gene="ECs4723"
/note="similar to WECD_ECOLI gi|1790225 percent identity 98 in 181 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38146.1"
/db_xref="GI:13364199"
/translation="MQAKAASNTGELDALQIQGFSLVEGEVDLALPVNNA SDSGAVV AQETDIPALQASAPQSRFPWPYAPDASRFYQWIEANVKTGFPHQCLIRAA SGIRGYVSLRNLNATDARIGLIAGRGAGAEIMQALNWAYARGKTTLRVATQMNGTA ALKRYIQSGANVESTAYWLVR"
653..1783
/gene="ECs4724"
653..1783
/gene="ECs4724"
/note="similar to WECE_ECOLI gi|2367285 percent identity 99 in 376 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative regulator"
/protein_id="BAB38147.1"
/db_xref="GI:13364200"
/translation="MIFPNAPPVVGTELDYMQSGKLGCGDGGFTRRCQWLQRF GSKAVLTSCSTASLEAALLDIQGDVIMPSYTFVSTANAFVLRGAKIVFVDRP DTWNIDETLIEAATIDKTRVIVPVHYAGVACEDTIALAKKHLNLFVDEDAQVWST YKGRALGTIGHICGFSFETKNYTAGGEGATLINDKALIEPAEIIREKGTNRGOFPR QVDKTYTWIDIGSSYLSLQAYLWAEADRLNQRLALWQNYDALAFELAKGR IELPSIPDGCLQNAHFYIKLRDIGNRSALINFLKEAEIMAVFHYIPLHGCPAGERFG EFHGDRYTTKSESRLLRLPLFNLSVNVNQTVIATLLNYS"
1785..3035
/gene="ECs4725"
1785..3035
/gene="ECs4725"
/note="similar to WZXE_ECOLI gi|1790227 percent identity 98 in 416 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative cytochrome"
/protein_id="BAB38148.1"
/db_xref="GI:13364201"

```
/translation="MSLAKASLWTAASLTAVKIGAGLVGKLLAVSFGPAGLGLAANPR
QITVLVLAGAGIENGVTKYVAQVHDNPOQLRRVGTSTSAVLGFTSLMALVFLAA
APISLGLFNGHYQGLVRLVAVOMGIANGNLLALMKGPRDAAGNALSILVGLIGV
LAVYSYRUGGVEGALLGLALPALVIPAAMLIKRGATPLSYLKPSWDNGLAGOLS
KPTLTLITSVLPVAYIMWRKLLAAQYSDWVGIGWGVSSIADVLOFQITASPSVYL
LPTLSRLTEKRITREVKYSLKXPLVAPVAAASFTVLLRDFAIWLLSNKFTANRDLF
AQVQGVDLKVAIYVFGVILVAKASRFLYIAEVSQFTLLMVFAHWLIPHTHGAUGAQ
AMATTVIVFSSCCGCVFLLMRRRA"
3032..4111
/gene="ECs4726"
CDS
3032..4111
/gene="ECs4726"
/notes="N-terminal part (1-62 in 359 aa) is similar to
YIFM_1 ECOLI gi|2367286 percent identity 96 in 62 aa,
C-terminal part (156-359 in 359 aa) is similar to
YIFM_2 ECOLI gi|2367287 percent identity 96 in 204 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38149.1"
/db_xref="GI:13364202"
/translation="MTVLHVLGSDIPHHNRTVLRFNDALAAATSEHAREFMVVGKDD
GUSDSPALSVOFFPFGKSLABAVIAKAKARQORFFPHGQFNPTLWLALISGIRPS
QYWHYLGADVELSLGRLKLYFLPRLRAKRVCFVATRGDLSFFAKTHPKVRGEL
LYPTMTDASLTANDROREKGMILVNGSDRSNEHIAALRAVHQFGDGVKVVVP
MGVPPNEAYIREVRQAGLELSEENLOVLSEKLEFDAYLLRQCGLGYEIRARQG
ICTLCLLIQAGIPCVLRNPNPQWQTEQHLPLVLTDDLDNEDIVRAQRQLASVDEN
TIAFFSPNYLQGWQRALAPAGEVA"
4108..5460
/gene="ECs4727"
CDS
4108..5460
/gene="ECs4727"
/notes="similar to WECF ECOLI gi|2367288 percent identity
99 in 450 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="TDP-FucANAC:lipidii transferase"
/protein_id="BAB38150.1"
/db_xref="GI:13364203"
/translation="MSLIQFSLGVVWMLLCTLIATLTWFFRFRVRENVPFSLPL
LFFFGFTSVLVRFDVGVAPPILLQALLSAGCFYAVYVYVYKTRLRKRVADVR
RLFTNVRVETNLVGTALVSQVGIFFMHNGFLFLRLNSYSQIFSSVSGVALKR
FYFFIPAMLVYFVRQSKAWLFLVSTVAFGLTYMVGCTRANIIIAFAPLFLFG
IIRGMSLWMLAAVGLGVFWLALRYGNVSGDEAFYFLYLRDTRDFSPWENLA
LLIQNDNTDFQOLAPIVDFVFIPIPSWLPGRPSMVLNSANYTWEVLNNHSGLAS
PTLIGSLVWVGALFPLCAIVVGLI IKWFDWLBELGNRETRNRYKAAIILHSFPGALF
NMIVLAREGLDSFSVRVFFIVVFGACLMIAKLLWLFPESAGLHKTRKSLRQVEG"
5463..6203
/gene="ECs4728"
CDS
5463..6203
/gene="ECs4728"
/notes="similar to WECG ECOLI gi|2367289 percent identity
98 in 246 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="probable UDP-N-acetyl-D-mannosaminuronic acid
transferase"
/protein_id="BAB38151.1"
/db_xref="GI:13364204"
/translation="MNNNTAETVTLRGLIGHWDQHALDYLFDGOLKQCTLVAT
NASKMTIBENAEVRELINAAEFKADGISVVRKYPQARQVSRVAGADLHEELMA
RAGKETTFLVGGPEVLEQAEKLRQNMVNIIVGSDGYFKEPQKQALFERIHAG
AQIVTVAMGSPKQEIFMRDCRLVHPDALYMGVGGTYDFTGHVKRAPKIMQTLGLEWL
YRLSQPSRIKQLRLLRYLRWHYTGNL"
6394..7779
/gene="ECs4729"
CDS
6394..7779
/gene="ECs4729"
/notes="similar to YIFK_ECOLI gi|2367290 percent identity
```

```
98 in 461 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative amino acid/amine transport protein"
/protein_id="BAB38152.1"
/db_xref="GI:13364205"
/translation="MADNKPPELQRLQEARHIELIAGLTGTVGLFWGAASLTLRWAGPS
VLLAYIIAGLVFFPFRMSGEMLFLEVTGTSFAVYAHRYMSPFGYLTATMSYFMWMA
VGEISITAGIVYQVQFPENMAQMIIPALIALVALANLAARVYGIETEFMAFIKVTY
IIVMIVIGLVIPFGFGNGQSIGFSLNTEHGGFFAGGKGFALTALCIVVASVQGVLE"
Query Match 95.5%; Score 1146.6; DB 1; Length 318703;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1181; Conservative 0; Mismatches 19; Indels 3; Gaps 2;
QY 1 AGAATAATATGTGAGATCGCACCCGCCATCGAATGTGCCAGTATATAGCGCTTTACGCCAC 60
DB 38026 AGAATAATATGTGAGATCGCACCCGCCATCGAATGTGCCAGTATATAGCGCTTTACGCCAC 37967
QY 61 GGACCGGCTGAACCTCTCTGTCGAGAAATGCCGCCAGATCATCAACATAATCATTAAG 120
DB 37966 GGACCGGCTGGAACCTCTCTGTCGAGAAATGCCGCCAGATCATCAACATAATCATTAAG 37907
QY 121 CGATTAAATGCCCCGAGATCGGATCGCTAACAGGCGACCGGAACGTCCTCGCCCGCA 180
DB 37906 CGATTAAATGCCCCGAGATCGGATCGCTAACAGGCGACCGGAACGTCCTCGCCCGCA 37847
QY 181 TGTGCGATGATTAGACATCAAAACCCCAATGGAACAGGTCTATAGGCCAGTTCGGCATAT 240
DB 37846 TGTGCGATGATTAGACATCAAAACCCCAATGGAACAGGTCTATAGGCCAGTTCGGCATAT 37787
QY 241 TTTTACTAGTCTCAATACGCGCGGCGAGATGACTACCAACCGGTCATGGTGTGTGCG 300
DB 37786 TTTTACTAGTCTCAATACGCGCGGCGAGATGACTACCAACCGGTCATGGTGTGTGCG 37727
QY 301 CGAAAAACGGAACAGCGCACCGGAATGTCTCACACAGTAACTCTGCTTCATCAGC 360
DB 37726 CGAAAAACGGAACAGCGCACCGGAATGTCTCACACAGTAACTCTGCTTCATCAGC 37667
QY 361 TCACGCCAGAAATCAGTCAGCGGTCCTCATGTAAGCAAGCAACGCGTTTCTCTTGT 420
DB 37666 TCACGCCAGAAATCAGTCAGCGGTCCTCATGTAAGCAAGCAACGCGTTTCTCTTGT 37607
QY 421 TCCAGTCTTTTGTGCTGTAACATCGGTAATCTCGCTCTTAAACACAGTAAATCGT 480
DB 37606 TCCAGTCTTTTGTGCTGTAACATCGGTAATCTCGCTCTTAAACACAGTAAATCGT 37547
QY 481 TTTTATTAGCGTGCCTGACACACGCTGCGACAGTATGTTGGCACAAAAATAGAC 540
DB 37546 TTTTATTAGCGTGCCTGACACACGCTGCGACAGTATGTTGGCACAAAAATAGAC 37487
QY 541 ACACCGGAGTTTCATGACTTAGAATGTTGGTTCCTTACCTGCTGACATGATCAT 600
DB 37486 ACACCGGAGTTTCATGACTTAGAATGTTGGTTCCTTACCTGCTGACATGATCAT 37427
QY 601 TTTAAGCTGTGCGCAGGCTCTGTTGCAATCAACACTATGACCACTCGCTCAACACGG 660
DB 37426 TTTAAGCTGTGCGCAGGCTCTGTTGCAATCAACACTATGACCACTCGCTCAACACGG 37367
QY 661 TTAT--CGGCGCGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATATTGT 718
DB 37366 TTATCGCGCGCGTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATATTGT 37307
QY 719 GCTGTTGGCGTGGGTT--GGAGCGCTATTTTCCGCTCAGTGAATTCGCTTCAAGTGT 777
DB 37306 GCTGTTGGCGTGGGTTGGGGAACGCTATTTTCCGCTCAGTGAATTCGCTTCAAGTGT 37247
QY 778 GAAGTGGCGCAGCGCGCTTACTTGTGTTGGTGGGAATCCAGCAGTGGCGCGCGCTGG 837
DB 37246 GAAGTGGCGCAGCGCGCTTACTTGTGTTGGTGGGAATCCAGCAGTGGCGCGCGCTGG 37187
QY 838 TCCAAATTGACCTTAAATCGTGGGCTCTACTCAATCGCTGCGACATTTGTTCCAGCGCG 897
```

Db 37186 CGCAATTGATCTTAATCGCTGGCTCTACTCAATCGCGTGCAGATTTGTTCCAGCGCGC 37127
QY 898 AGTTTGTGGAATCTCACCACATCCCAAAAGTATTGTGTTCTGCGGCGGTATTTCGCGCA 957
Db 37126 AGTTTGTGGAATCTCACCACATCCCAAAAGTATTGTGTTCTGCGGCGGTATTTCGCGCA 37067
QY 958 ATTATCATGCGCGCAACGCGCACTGATCGAGTATATCGTCTCGGCTCACCACATAT 1017
Db 37066 ATTATCATGCGCGCAACGCGCACTGATCGAGTATATCGTCTCGGCTCACCACATAT 37007
QY 1018 TGTGTTGCGATATTATTGTGATGATCGGTTAGCCACCTTCTCTCAACGGGATTGCTCTATG 1077
Db 37006 TGTGTTGCGATATTATTGTGATGATCGGTTAGCCACCTTCTCTCAACGGGATTGCTCTATG 36947
QY 1078 GATTAAAGACCAAGACAGATGAAGCGCTGAATAAGATTTCGCTCGTGTGTTATGCT 1137
Db 36946 GATTAAAGACCAAGACAGATGAAGCGCTGAATAAGATTTCGCTCGTGTGTTATGCT 36887
QY 1138 GTGGGAGCGCTGTTAGCATCGCGGAGCATGCGTGAATAATATGTCGATCGCGGCTA 1197
Db 36886 GTGGGAGCGCTGTTAGCATCGCGGAGCATGCGTGAATAATATGTCGATCGCGGCTA 36827
QY 1198 AAC 1200
Db 36826 AAC 36824

RESULT 9
AE016769/c
LOCUS
DEFINITION Escherichia coli CFT073 section 15 of 18 of the complete genome.
ACCESSION AE016769 AE014075
VERSION AE016769.1 GI:26110701
KEYWORDS
SOURCE Escherichia coli CFT073
ORGANISM Escherichia coli CFT073
REFERENCE
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1. (bases 1 to 301566)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Rossch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.P., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Domeneberg,M.S. and Blattner,F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157
2. (bases 1 to 301566)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Rossch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.P., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Domeneberg,M.S. and Blattner,F.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..301566
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(72..1100)
/gene="waaW"
/locus_tag="c4450"
complement(72..1100)
/gene="waaW"
/locus_tag="c4450"
/note="Residues 1 to 341 of 342 are 98.24 pct identical to
residues 1 to 341 of 341 from GenPept.129 :
sbl/AAC9673.1 (AF019746) UDP-galactose (galactosyl) LPS
alpha1,2-galactosyltransferase WaaW [Escherichia coli]"
/codon_start=1

/transl_table=11
/product="UDP-galactose (galactosyl) LPS
alpha1,2-galactosyltransferase"
/protein_id="AA082886.1"
/db_xref="GI:26110702"
/translation="MDLAESITEVAVSGEIANTRVNLAYGIDRNFLEGAASVMSQ
VMNPLAVKFLFDYIDEDYLQVNAFTSKNVEVRIYKVSFAFIDIFSLQW
SYATFLVAFQYLSEIENLLYDADVICKSLAGLLDINFQDKFAAVIKDVPFMQ
EKPAKRLAIEGLPGNYAGVYQLEAWKNDPMKAIAMLASDPQHTYKCLDDI
LNLFFHCIFISGDYDCFYIDYELKNSDEYKRTITDDTKLIHYGVYKPNOWT
NYPQKYENAYQVSCWNVAFIPATNEKQYQVKYCHAKNGDTENAFYIFKFLNK
YKRLFGQ"
complement(1126..1818)
/gene="rfay"
/locus_tag="c4451"
complement(1126..1818)
/gene="rfay"
/locus_tag="c4451"
/functions="enzyme; Macromolecule metabolism;
Lipopolysaccharide"
/note="Escherichia coli K-12 ortholog: b3625; Escherichia
coli O157:H7 ortholog: z5052"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide core biosynthesis protein
rfay"
/protein_id="AA082887.1"
/db_xref="GI:26110703"
/translation="MITSIRYGRGSPFYKDNKNKYKEIFEILLAYNFKTVKLRNIDD
TKSLIDTKYRVKVFAPKTRNERFLKSPVKGYQNLIVETDVRVSAGLTPFND
FYFLAEKFIYNAVFIMLEYVEVGLNDMPFIPEDKAEIKASKEKHALNLSGD
PHRGFTVSDGVRIIIDLSKSTAEKARDRLMERHLGIANEIKDYGYSVIYRTK
LRPKFKLKGKA"
complement(1828..2823)
/gene="rfaj"
/locus_tag="c4452"
complement(1828..2823)
/gene="rfaj"
/locus_tag="c4452"
/EC_number="2.4.1.58"
/functions="enzyme; Surface polysaccharides and antigens"
/note="Escherichia coli K-12 ortholog: b3626; Escherichia
coli O157:H7 ortholog: z5051"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide 1,2-glucosyltransferase"
/protein_id="AA082888.1"
/db_xref="GI:26110704"
/translation="MNEFIKERPSYLDNKNVPELVUSYGDKNFLYAGVSISV
LINSDFNVFVFTDYDDYILKSNFETAKQFNSTIILIDPKIFADLPISQFWSY
ATFVRLSPFYLSESITLLYLDADVCKGSLKPTKLIIFKDFEFAVDPNDSTQAAC
AKRLNIPGMNGRYFNAGVIYNLKWHEANLPYLLTLLRGETKYGSLYDQDALNI
AFNNANIYLAQDEDTTYLKNLHLSHRKYQQTITDKTVLIHYTGITKPMHWSAGYP
SAGVFNAREQSPWKYPLKEARTVAEMQKVKHLFAHGEYIKGITSLIKYLKKK"
complement(2840..3856)
/gene="rfal"
/locus_tag="c4453"
complement(2840..3856)
/gene="rfal"
/locus_tag="c4453"
/EC_number="2.4.1.44"
/functions="enzyme; Surface polysaccharides and antigens"
/note="Escherichia coli K-12 ortholog: b3627; Escherichia
coli O157:H7 ortholog: z5053"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide 1,3-galactosyltransferase"
/protein_id="AA082889.1"
/db_xref="GI:26110705"
/translation="MSAHVFNQEMINKTIIEDERPAVSASSPHVAYGIDONELFGC
GVSTVLLHNNDSVSVFHVIFDDIEPAIQLSLAKSYHTICIHVNCERLKAIP
TKNWSIATMFRVIADYFDQDKLILYDADIACQGNLKPITIMDLANNVAUVTER
DANWLSRAQSLQCNLEKGYFNISVLLINTLAWAQESVSAKAMSLADKAVVSLTY

gene
MQDILNLLGLKVFIDAKYNTQFSLNLYELKKGFIQPCINDETSLIHYVGTPEWHVW
AGVPSAQPFKIGKEASPKWNEPLMRPMNSVARYCAKHNFKQKPNIGWYIYFYL
KIIL
complement (3872..4678)
/gene="rfap"
/locus_tag="c4454"
complement (3872..4678)
/gene="rfap"
/locus_tag="c4454"
function="enzyme; Macromolecule metabolism;
Lipopolysaccharide"
/note="Escherichia coli K-12 ortholog: b3630; Escherichia
coli O157:H7 ortholog: z5054"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide core biosynthesis protein
rfap"
/protein_id="AAN82890.1"
/db_xref="GI:26110706"
/translation="MWKVELKEPEATLWRGKDFEEVKILQGEVFELETRRLTAPL
MAGSYFLKWHRTGLTKIILKNLSLMPVLGADREWNATHRLRDVGVDVIMYGAFFE
KINLETRFTITLEDLPTISLEDYCADWATNPDPVRKMLIKRYATVWRDYMHAAG
INHRDCYIHFLLHLPFGSKBEELKISVIDLFRALQTRVPRWRDXDLGLYFSSNN
IGLTQDRIWRFKWVVSAPLKDILKQEQGLLSQAEAKATKIRETIRKSL"
complement (4662..5786)
/gene="rfag"
/locus_tag="c4455"
complement (4662..5786)
/gene="rfag"
/locus_tag="c4455"
function="enzyme; Macromolecule metabolism;
Lipopolysaccharide"
/note="Glucosyltransferase I; Escherichia coli K-12
ortholog: b3631; Escherichia coli O157:H7 ortholog: z5055"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide core biosynthesis protein
rfag"
/protein_id="AAN82891.1"
/db_xref="GI:26110707"
/translation="MIVAFCLYKYPFGGLQDFWRFACTVAARGHVRVYVTSWGE
CDVFEILKVPKSTHNGRAEYPAWQKLEHREHPVDRVVGKMPGLDVIYAADVC
YAEKVAQEGFFRYLTSRYHAAFERATFEQKPTQLMLTDKQIADFOKHVCTAAE
RHILPPGYPDPYTSQFPANSEIRFRKNGITEQYLLIQVDSDFTRKGVDSIAL
ASLPDSRLHTLLYVGGQKPKFEALAKGVRSNVHFFSGRNDVSELMAAADLLH
PAYQBAAGVLLLEATAGLPLVTTAVCGYAHYVDANGGAIAPPPQETLNEILKA
LTQSSLRQAWENASHYADTDQLYSLPEKADIITGSLDG"
complement (5783..6841)
/gene="rfag"
/locus_tag="c4456"
complement (5783..6841)
/gene="rfag"
/locus_tag="c4456"
/BC_number="2..-..-"
function="enzyme; Macromolecule metabolism;
Lipopolysaccharide"
/note="Escherichia coli K-12 ortholog: b3632; Escherichia
coli O157:H7 ortholog: z5056"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide core biosynthesis glycosyl
transferase rfag"
/protein_id="AAN82892.1"
/db_xref="GI:26110708"
/translation="WDKPFRIILIKMFHGMMLLTFVSSLKKNPYDAKIDVLLYQ
DTIPILSENFENALYGNKKAKASEKIANFFYLIKVRANKYDLVNLTDQMWAI
LVRLLNARVIGIDYVHRRQSAFWRNSFHLVPLQGNVNSNLSVLPGLSILVKQT
TWSYPASWKRWRRLDHAGVQNVVITOPTARGIKFCWONAKFSAVIDALHARGVEV
VLTSGPDKDLACVNEIAGGCTPVTALAGKVFPELGLIDHAQLFGVDSAPAHIA
AAVANTPLISLFGATDHFWRPWSNNMQFWAGYREMPEDQDRNEMVLSVIPAAD
VIAAVDKLLPSSITGSL"
7254..8531
/gene="kda"

CDS
/locus_tag="c4457"
7254..8531
/gene="kda"
/locus_tag="c4457"
/EC_number="2..-..-"
/function="enzyme; Surface polysaccharides and antigens"
/note="Escherichia coli K-12 ortholog: b3633; Escherichia
coli O157:H7 ortholog: z5057"
/codon_start=1
/transl_table=11
/product="3-deoxy-D-manno-octulosonic-acid transferase"

Query Match 94.0%; Score 1128.2; DB 1; Length 301566;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1176; Conservative 0; Mismatches 23; Indels 4; Gaps 3;
QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
DB 281750 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 281691
QY 61 GGACCGGCTGAACCTCTGTGTGCAGAAATGCCGCCAGATCATCAACATAATCATTTAAAG 120
DB 281690 GGACCGGCTGAACCTCTGTGTGCAGAAATGCCGCCAGATCATCAACATAATCATTTAAAG 281631
QY 121 CGATTAAATGCCCGAGATCGGATCGGCTAAACAGGCGACCGGAACGTCCCTGCCCGCA 180
DB 281630 CGATTAAATGCCCGAGATGTGGATCGGCTAAACAGGCGACCGGAACGTCCCTGCCCGCA 281571
QY 181 TGTGTCGATGATTAAAGACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCGGCATAT 240
DB 281570 TGTGTCGATGATTAAACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCGGCATAT 281511
QY 241 TTTAGTGTCTCTCAATACCGCCCGGACAGATGATGATACACAGTAACTCTGTTCAATACGC 300
DB 281510 TTTAGTGTCTCTCAATACCGCCCGGACAGATGATGATACACAGTAACTCTGTTCAATACGC 281451
QY 301 CGAAAAACGACAAAGCGCACCGGAATGTCTATCCACACAGTAACTCTGTTCAATACGC 360
DB 281450 CGAAAAACGACAAAGCGGAATGTCTATCCACACAGTAACTCTGTTCAATACGC 281391
QY 361 TCAGCCAGAAATCAGTCAGCGGTCCTCATGTTAAAGCAGCAACCGGTTTCTCTTGT 420
DB 281390 TCAGCCAGAAATCAGTCAGCGGTCCTCATGTTAAAGCAGCAACCGGTTTCTCTTGT 281331
QY 421 TCCAGTCTTTTGTCTGTAACATCGGTAATCTGCTCTTTAAACACAGTAAATCGT 480
DB 281330 TCCAGTCTTTTGTCTGTAACATCGGTAATCTGCTCTTTAAACACAGTAAATCGT 281271
QY 481 TTTTGTAGCGTCTGACACACAGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 540
DB 281270 TTTTGTAGCGTCTGACACACAGCTGCGACAGTAGCGTATTGTGGCACAATAATAGAC 281211
QY 541 ACACCGGAGTTTCATGACCTTAGAATGTGTTTGCCTACTGCTGACATCGATCAT 600
DB 281210 ACACCGGAGTTTCATGACCTTAGAATGTGTTTGCCTACTGCTGACATCGATCAT 281151
QY 601 TTTAAGCTGTGCGCAGCTCTGTGCAATCAACATATGACCACTCGCTCAACACCG 660
DB 281150 TTTAAGCTGTGCGCAGCTCTGTGCGCAATCAACATATGACCACTCGCTCAACACCG 281091
QY 661 TTAT--CCGCGCGGTGGCTATTGTGGGCTTACAGCCGACTGGCGATTCTATTGT 718
DB 281090 TTATCGCGCGCGGTGGGCTATTGTGGGCTTACAGCCGACTGGCGATTCTATTGT 281031
QY 719 GCTGTTGGCGTGGGTT--GGAGCGCTATTTTCCGCTAGTGAATGCGTTGAAGTGT 777
DB 281030 GCTGTTGGCGTGGGTTGGGAGCGCTATTTTCCGCTAGTGAATGCGTTGAAGTGT 280971
QY 778 GAAGTGGCAGCGCGGCTACTTGTGCTGGGAATCCAGCAGTGGCGCGCGCTGG 837
DB 280970 GAAGTGGCAGCGCGAGCTTACTTGTGCTGGGAATCCAGCAGTGGCGCGCGCTGG 280911
QY 838 TGCATTTGACCTTAAATCGCTGGCCTTACTCAATCGCTGACATTTGTTCCAGCGCGC 897

```
Db 280910 CGCAATTGATCTTAAATCGCTGGCTCCACTCAATCGCTGCACATTTGTCACGCGCGC 280851
QY 898 AGTTTGTGATCTCACCATCCCAAGATTTGTTCTGCGCGCGTATTTCGCA 957
Db 280850 AGTTTGTGATCTCACCATCCCAAGATTTGTTCTGCGCGCGTATTTCGCA 280791
QY 958 ATTATCATCGCGCAACAGCGCACTGATCGATATATCGTCTGCGCGTACCACTAT 1017
Db 280790 ATTATCATCGCGCAACAGCGCACTGATCGATATATCGTCTGCGCGTACCACTAT 280731
QY 1018 TGTGTCGATATATGATGATCGGTAGCCACCTTCTCAACGGATGCTCTATG 1077
Db 280730 TGTGTCGATATATGATGATCGGTAGCCACCTTCTCAACGGATGCTCTATG 280671
QY 1078 GATTAAGACCAACAGCAGATGAAGCGCTGAAATAAGATTTTCGCGCTGTTTATGCT 1137
Db 280670 GATTAAGACCAACAGCAGATGAAGCGCTGAAATAAGATTTTCGCGCTGTTTATGCT 280611
QY 1138 GTTGGAGCGCTGTAGCATCGCGGAGCATCGGTGAAATAATATGTCGATCGCGCGTA 1197
Db 280610 GTTGGAGCGCTGTAGCATCGCGGAGCATCGGTG-AAAATAATGTCGATCGCGCGTA 280552
QY 1198 AAC 1200
Db 280551 AAC 280549

RESULT 10
AL627278 258050 bp DNA linear BCT 04-JUL-2003
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome, segment 14/20.
ACCESSION AL627278 AU513382
VERSION AL627278.1 GI:16504263
SOURCE Salmonella enterica subsp. enterica serovar Typhi
ORGANISM Salmonella enterica subsp. enterica serovar Typhi
Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 258050)
Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,
Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Conerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,
Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
Tagatz, K., Krögh, A., Larsen, T.S., Leather, S., Moule, S., O'Garra, P.,
Perry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrell, B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
Nature 413 (6958), 848-852 (2001)
JOURNAL MEDLINE
21534947
PUBMED 11677608
REFERENCE 2 (bases 1 to 258050)
Farkhill, J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/s_typhi/).
FEATURES
source
1. 258050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
69..1544
gene
```

```
/gene="tolC"
/notes="synonym: STY3364"
69..1544
/gene="tolC"
/notes="Similar to Salmonella enteritidis outer membrane
protein TolC precursor tolC SW:TOIC_SALEN (Q54001) (491
aa) fasta scores: E(): 0, 99.0% id in 491 aa, and to
Escherichia coli outer membrane protein tolC precursor
tolC or mcb or muka or refi SW:TOIC_ECOLI (P02930) (495
aa) fasta scores: E(): 0, 89.7% id in 495 aa
Orthologue of E. coli tolC (TOIC_ECOLI); Fasta hit to
TOIC_ECOLI (495 aa), 90% identity in 495 aa overlap"
/codon_start=1
/transl_table=11
/product="outer membrane protein TolC precursor"
/protein_id="CAD07712.1"
/db_xref="GI:16504264"
/db_xref="GOA:O823N8"
/db_xref="SPREMBL:O823N8"
/transl_table="MOMKLLPILIGLSGFSLTQAEHLQVYQARLSNPELRKS
AADRAPEKINERASPLLPOLGADITYNSGRDANGINSNETSASLOITLFDL
SKWGLTLQKAAGIQDVTYTDQOTLINTANAYFKVLNADVLSTYQAKQEAIRQ
LDQTFQFNVLVAITDQVNAQYDTLVANEVTARNLDNAVEELQVGTGYPLLA
SLNVEHFKTDKPAVNALLKEAENRLSLQARLSQDLAREQIRQADQHLPTLNLT
STGISTSYSGKTSAGYDDSNMGNKIGLNFSLPLYOGMVNSVKQAQYNFVGAS
EQLESAHRSYVQTVRSFNNINASI SINAYKQAVUSQSSLDAMEAGYSGVPTIVD
VLDAITTYLDAKQQLANARYTILINQLNKYALGTILNEQDLLALNSTLQKPIFSPES
VAPETPEQDAADGYNASAPVQPTAARANSNNGNPFPH"
1757..2428
/gene="STY3365"
1757..2428
/gene="STY3365"
/notes="Orthologue of E. coli ygiB (YGIB_ECOLI); Fasta hit
to YGIB_ECOLI (234 aa), 97% identity in 223 aa overlap.
Contains a possible N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="possible lipoprotein"
/protein_id="CAD07713.1"
/db_xref="GI:16504265"
/db_xref="SPREMBL:Q8XGZ1"
/transl_table="MKTKSIHRAFSRKSWSARHLTPVALAVTAVMLACSEKSDTV
SLYNQADCSAANFGSAECTATNNALKEARTAPKYATREDCAVAFEGGCGQAPA
QAGNAPNQAAQSGSGFWMLMAGYMRGLMGGAQFPAQPLFSKRPAPAYGKY
TDAAGKNYGAQPGRTWTVKTAAPKATTTVTTRGFGESVAKQSTWQSAAGTST
RSMG3"
1835..1867
/misc_feature
/gene="STY3365"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
2434..3597
/gene="STY3366"
2434..3597
/gene="STY3366"
/notes="Fasta hit to YJFC_ECOLI (387 aa), 50% identity in
392 aa overlap
Orthologue of E. coli ygiC (YGIC_ECOLI); Fasta hit to
YGIC_ECOLI (386 aa), 94% identity in 386 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD07714.1"
/db_xref="GI:16504266"
/db_xref="SPREMBL:Q8XFF6"
/transl_table="MERVSITERPDWRDKATEYGFNHTMYGPFYWCEDAYKLTAAQ
VEKLEDVABLHOMCLKXVERVIADELMTKFRPKHTWGFVQSQPQPSLYSLRL
LAWGIGEPKLENNADTPTSLYAAAFWLEDOINAGNLPEGSDQFNLSQEKLE
FPAERLQYGFOLLHLCCTCDTVEDRGTIOYLQCAABAEIATEFYINDDIGLGEKQ
FTDLQDVIANLFLKYEFWFLREMPSTKLEDAGVNLPEPAKSLISNKALPLNEM
FPDFNLLPAYFAEDHPMDKIVKPIFERGAGNVSIENGKTIIESVGPVGEEMI
VQFPIPLFGDSYTLIGSLINDQFAGIGREDRALITDLSRFYPHIFVEG"
complement (3659..4489)
/gene="STY3367"
```


CDS complement (3659. .4489)

```

complement(3659. .4489)
/notes="Orthologue of E. coli ygiD (YGI_D_ECOLI); Fasta hit to YGI_D_ECOLI (271 aa), 87% identity in 266 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD07715.1"
/db_xref="GI:16504267"
/db_xref="GOA:Q8Z3N7"
/db_xref="SP:PREMBL:Q8Z3N7"
/translations="MYMLKTLCSKDNIMSLTCTMPALFLGHGSPMVLDDNDYTRAVR
RLGEALPRQAIIVVSAKHVYETGCTVAMERPOYLHDQFPGPOALVDMHYPAEGSPAL
AQRLVELLAPGVVALDKAMRGFDHGSWGVLKNFENADIPNVLQSLVDSTKPAAMHREM
GRKLATLDEPVLVAGNVNHNLRKTVARWGHNDITPIWMAASFDFVKANITVPGCPVEYQ
HPLVNYLQIEGGALSNPTEHFLLPYVLGADWDKKEPTITPVDGIEWGSISMLSVQVG
"
4588. .5361
/genes="STY3368"
4588. .5361
/genes="STY3368"
/notes="Orthologue of E. coli ygiE (YGI_E_ECOLI); Fasta hit to YGI_E_ECOLI (257 aa), 93% identity in 257 aa overlap.
Contains multiple possible membrane spanning hydrophobic
domains"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAD07716.1"
/db_xref="GI:16504268"
/db_xref="GOA:Q8XGR4"
/db_xref="SWISS-PROT:O8XGR4"
/translations="MSVPELITLLAGATFTFAGVLGQKPSNRVLAFSLGFAAGIM
LLISLEMLPAALDTGMSPFVLGYGMFLIGLVGGLDRLAPHADHLPQKQSQPLP
GSKIRKATLTLGYSUNTPEGEATFVATSSNLELIGLIALAHVNIPEGLAVAGPV
YAATGSKRTAIFWAGISGMAEILGGVLAWSLVSPIVNAALMAAVAGIMVALSVSD
ELMPLAKEDINPNPSYGLVCGMSINGLSLVLTIGTG"
6004. .7800
/genes="STY3370"
/notes="synonym: asst"
6004. .7800
/genes="STY3370"
/notes="Similar to Klebsiella pneumoniae arylsulfate
sulfotransferase asst TR:P97036 (EMBL:U32616) (598 aa)
fasta scores: E(): 0, 86.8% id in 598 aa"
/codon_start=1
/transl_table=11
/product="Probable arylsulfate sulfotransferase"
/protein_id="CAD07717.1"
/db_xref="GI:16504269"
/db_xref="GOA:Q8Z3N6"
/db_xref="SP:PREMBL:Q8Z3N6"
/translations="MFEDYRKTLIAGAVALTICGLTAAGTFAAGFQAPQAPAGKLGAVV
DPYGNALITALVELDHSILSDVKTVHGKGEKGPVTVTGKBSLETVDGIPIFGLYG
KFAINVVEYKENGKAMKDDVYVQTSALVNHVMDNRISIDLCQKVKIVKAPGEDRLY
LVNTHTPQGAERPMHGEKDKNAGLIDGAPGAGALPPDIATPYTFVDVTQGEIKWMLD
QDTFYNGHDNNKRGYLMGIRETPGRTFTAVQGHWFVEFDMGQILADHKLPRGFLD
ASHESIETVANGTVLLVRGDEYRKDGIVHEITRQIETVQSGRWVDWMLTAYDAK
MRDALLAGDAGVGVNVDLAHAGQAQLEPPTPGDGLVCGRWVHNLSITAYDAK
DSITILSRFHQIVKIGRQKVTWVILPASKGNKOLASKLLPKVDHGKPTKIDENGK
CKDPTDPTQHTWATLSSKQTLFVLDGNGRGLRQPALPTWYKYSRFVEYKIDKEKGT
VQQWVEYKRGYDIFSPITSVVEYQKDDRTMFGFGGGINLPDVGKPTVGLKLNELDKY
TKVKVEDVLSDKPNQHTALLVHPTQMFK"
7820. .8491
/genes="dsba"
/notes="synonym: STY3371"
7820. .8491
/genes="dsba"
/notes="Similar to Enterobacter amnigenus disulfide
isomerase dsba TR:Q9XDR1 (EMBL:AF012826) (222 aa) fasta
scores: E(): 0, 90.1% id in 222 aa"
/codon_start=1
/transl_table=11

```



```

CDS
/locus_tag="t3219"
complement(4761..5726)
/locus_tag="t3219"
/note="corresponds to STV3481 from Accession AL513382:
Salmonella typhi Ctr18"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAO70755.1"
/db_xref="GI:29139188"
/translation="MSKQAGIGILALTAMCWGALPTAMKQVLEWEPSTIVFRL
MASIGALAVKRLPLRIKPKRWLVILAIATCGLPFNILFSSLSQVLSPTAQ
VIGLSPVMVASVFLIKERQGTQVIGALMLLSGLVFNFTSLIIFTRLDYTWG
VIFGGAAMVVSYGVAQVLLRLASQIILFLYLCTIALLLPLAKPMVIAQSDWQ
LACILFCGNTLVGVALAENARQAAQVSAIITLPLFLFSLLSWAPDFPAR
PMLNLLGLCAFWVAGAMYSAGHRIWGLRGKHEVTVVSQPSRSG"
/gene="rpsA"
/locus_tag="t3220"
complement(5856..6113)
/locus_tag="t3220"
CDS
complement(5856..6113)
/gene="rpsA"
/locus_tag="t3220"
/note="corresponds to STV3482 from Accession AL513382:
Salmonella typhi Ctr18"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L27"
/protein_id="AAO70756.1"
/db_xref="GI:29139189"
/translation="MAHKAGGSTRNGRDSKRLGVKRFGEAVLAGSIIVRQGTG
FHAGTVGGRDHTLFAKADGVKPKVKNKRYISVAE"
/gene="rplU"
/locus_tag="t3221"
complement(6133..6444)
/locus_tag="t3221"
CDS
complement(6133..6444)
/gene="rplU"
/locus_tag="t3221"
/note="corresponds to STV3483 from Accession AL513382:
Salmonella typhi Ctr18"
/codon_start=1
/transl_table=11
/product="50S ribosomal subunit protein L21"
/protein_id="AAO70757.1"
/db_xref="GI:29139190"
/translation="MTAVFSQSGKHVRSEGTQVLEKLDIATGETIEFAEVLMIANG
EYVKIGVPPVGGVKAIEVVAHGEKVKIVKFRKRRKVRKQGHQRFNTDVKLTGIS
A"
/locus_tag="t3222"
complement(6703..7674)
/locus_tag="t3222"
CDS
complement(6703..7674)
/gene="ispB"
/locus_tag="t3222"
/note="corresponds to STV3484 from Accession AL513382:
Salmonella typhi Ctr18"
/codon_start=1
/transl_table=11
/product="octaprenyl-diphosphate synthase"
/protein_id="AAO70758.1"
/db_xref="GI:29139191"
/translation="MNLKXNELTAQDMAGVNAITLQNLSPVOLNOLGYVLIISGGG
KRIRPMVLAABAVCYQGNHVTIAALIEFTLTLHDDVVDVDSMDRERKATANAA
FNAASVIVGFYITRAFQNMVSLGSLKVLSEAVNVIAEGEVLQMLMNVNDEITE
ENTNRVLYSTADLFKAAQCSGLAGCTPEQKGLQDYGRYLGFAPQLIDLLDIYSA
DGHLGNVGDLINEKPTLPLHAMRHGTPEQSAMITAIPOQNGRHLLSPVLEAMT
TCGSLWTRQRABEEADKAISALQILFDPFWREALIGLAHIAVQRDR"
/gene="nlp"
/locus_tag="t3223"
complement(7907..8194)
/locus_tag="t3223"
CDS
complement(7907..8194)
/gene="nlp"
/locus_tag="t3223"
/note="corresponds to STV3485 from Accession AL513382:

```

```

Salmonella typhi Ctr18"
/codon_start=1
/transl_table=11
/product="Ner-like regulatory protein"
/protein_id="AAO70759.1"
/db_xref="GI:29139192"
/translation="WESKLIDWHPADIAGLRKKGTSMAAESRRNGLSSTLANALTR"

Query Match      62.9%; Score 754.6; DB 1; Length 299991;
Best Local Similarity 78.6%; Pred. No. 1.8e-223;
Matches 927; Conservative 0; Mismatches 249; Indels 3; Gaps 2;

QY 1 AGAATAATGTGGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCTTTACGCCAC 60
DB 132183 AGGAACAGGGTGGCTATCGCCCCCCCCATTGAGTGCGCCAGGATATACGCTTACGCCAC 132242
QY 61 GGACCGGGTGAACCTCTCTGTCGAGAAATGCCGCCAGATCATCAACATAATCATTAAG 120
DB 132243 GACCCCGGCTCAATTTCTGCTGCCAGAACCGCGCTAAGTCTCAACATAATGTTAAAA 132302
QY 121 CGATTAAATGCCCGAGATGCGGATCGGCTAACAGGACCGGAACTCCCTGCCCGCGA 180
DB 132303 TGATCGACATGACCCCGATGCGGGTCCGATAACATCCGCCCGGAGCGCCCTTGACACGG 132362
QY 181 TGGTCGATGATTAGACATCAAAACCCCAATGGAACAGGTCTATAGGCCAGTTCGCGATAT 240
DB 132363 TGGTCGATGATGAATATCAAAACCGAGATGAAGAAGTCTATAGCCAGTTCAGCGTAT 132422
QY 241 TTATCGTAGCTCTCAATAGCCCGCGGAGATGACTACACCCCGGTATGGTGTGTGGCG 300
DB 132423 TTATACATCACTCAATCGGCCAGGACAAATGATGATCGTCCGATCGTTGGAGTCAATC 132482
QY 301 CGAAAACGGACAAAGCGACCGGAATGTCACACACCCAGTAAACTCTGCTTCAATCACGC 360
DB 132483 CGAAAACGGACAAAGCGACCGGAATGTTATCCAGCGGATAAACTCGGCTCTTCCGCG 132542
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGTGTAAGAAAGCAGCAAAACGCGTTTCTCTGTT 420
DB 132543 TGACGCCAGAAATCAGTCAGCGGTCCCATGTGTAAGAAAGCAGCAAAACGCGTTTCTCTGTT 132602
QY 421 TCCGAGTCTTTTGTGCTGTAACATCGGTAATCTGCCTCTTAAACACAGTAAATCGT 480
DB 132603 TCCGAGTCTTTTGTGCTGTAACATCGGTTTTCGGCTCTGTTTACAGTAAATCGT 132662
QY 481 TTTTGTAGCGTCCCTGACACAAACGCTCGCAGTAGCGTATTTGGGCACAAAAATAGAC 540
DB 132663 TTTTGTAGCGTCCCTGACACAAACGCTCGCAGTAGCGTATTTGGGCACAAAAACAGAC 132722
QY 541 ACACCGGAGTTCATCATGACCTTAGAATGGTGTTCCTACCTGCTGACATCGATCAT 600
DB 132723 AACCAAGGAGTTTCGCATGACCTTTGAATGGTGTTCGCTTATCTACTGACCTCAACTCT 132782
QY 601 TTTAACGCTGTCCCGAGGCTCTGTGTGCAATCAACACTATGACCACTCGCTCAACCCAGG 660
DB 132783 GTTGAGTCTTTCTCCGGGTTTCAGGCGCCATCAATACCATGACGACGCTAATCAACCATGG 132842
QY 661 TTATCCGGCGG--GTGGCGCTCTATTGCTGGGCTTCAGACCGGATCGGATTCATATGT 718
DB 132843 ATATCGTGGCGAGCGGCTTCTATCGCGGCTCCAGACCGGGTGGGGATACATATCGT 132902
QY 719 GCTGGTGTGGCGTGGGCTTGGG-ACGCTATTTTCCCGCTCAGTGTTCGTTTGAAGTGT 777
DB 132903 ACTGGTGTGGCGTTCGACATGGGTACGCTCTTTTCGGCTCGCTCATCGCTTTTGAATTC 132962
QY 778 GAAGTGGGAGCGCGGCTTACTTGAATTTGGCTGGGAATCCAGAGTGGCGCGCGCTGG 837
DB 132963 GAAATGGGCTGGCGGCTTATCTTATCTGGTGGGTATCCAGCAATGGCGCGCGCGAGG 133022
QY 838 TGCAATTGACCTTAATCGCTGGCTCTACTCAATCGGCTGACATTTGTTCCAGCGCGC 897
DB 133023 CGCTATCGATCTGCATCTCTCGCCACAGCGAATCGCGGGGTGGCTGTTCACACGGGC 133082
QY 898 AGTTTTGTGAATCTCACCAATCCCAAAAGATTGTGTTTCTGGCGGCGCTATTTCGCGCA 957

```

Db	133083	GATA-TTGTCAATCTAACCAATCCAAAGATTGTTCTTCTTSCGCCCTGTTCCGCA	133142
QY	958	ATTATCATATGCGCAACAGCGCAACTGATCAGCATATATGCTGCTGGCGCTACCACTAT	1017
Db	133143	ATTATCATATGCGCGAGCAACGCACTGGCGCACTACTTCTTCTGCGCTACCACTAT	133202
QY	1018	TGTGTCATATATTATGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG	1077
Db	133203	TGTGTCATATATTATGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG	133262
QY	1078	GATTAAAGACCAACAGCAGATGAAGCGCTCAATAAGATTTCGCTGCTGTTTATGCT	1137
Db	133263	GATTAAAGACCAACAGCAGATGAAGCGCTCAATAAGATTTCGCTGCTGTTTATGCT	133322
QY	1138	GTTGGAGCGCTGTTAGCATCGGCGAGGCGATGCTGAAA	1176
Db	133323	GGTAGGCGCTCTCTGCGCTGCGCAAGACACGCTGATA	133361
RESULT 12			
AE008884/c			
LOCUS	AE008884	21692 bp	DNA linear BCT 23-APR-2003
DEFINITION	Salmonella typhimurium LT2, section 188 of 220 of the complete genome.		
ACCESSION	AE008884	AE006468	
VERSION	AE008884.1	GI:16422503	
KEYWORDS	Salmonella typhimurium LT2		
SOURCE	Salmonella typhimurium LT2		
ORGANISM	Salmonella typhimurium LT2		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.		
AUTHORS	McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Gwral, N., Mulvaney, E., Ryan, E., Sun, H., Florea, B., Miller, W., Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.		
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium LT2		
JOURNAL	Nature 413 (6858), 852-856 (2001)		
MEDLINE	21534948		
PubMed	11677609		
REFERENCE	2 (bases 1 to 21692)		
AUTHORS	The Salmonella typhimurium Genome Sequencing Project		
CONSTRM	Direct Submission		
TITLE	Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA		
JOURNAL	COMMENT Supported by NIH grant 5U 01 AI43283		
COMMENT	Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/		
	EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/ , and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/		
	The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and RegionDB; http://kinich.cifn.unam.mx:8950/db/regulondb_intro.frameset		
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.		

FEATURES	source
gene	Location/Qualifiers 1. 21692 /organism="Salmonella typhimurium LT2" /mol_type="genomic DNA" /strain="LT2; SGSC 1412; ATCC 700720" /db_xref="ATCC:700720" /db_xref="taxon:99287" /note="LT2" /gene="cyaA" /note="synonym: STM3939" /gene="cyaA" /note="putative -35 signal for cyaA; RegulonDB:STMLTH004631" 136..151 /gene="cyaA" /note="putative binding site for CRP, RegulonDB:STMSIH000108" /bound_moiety="CRP" 137..145 /gene="cyaA" /note="putative -10 signal for cyaA; RegulonDB:STMLTH004631" 284..289 /gene="cyaA" /note="putative RBS for cyaA; RegulonDB:STMSIH003879" 297..2843 /gene="cyaA" /EC number="4.6.1.1" /note="adenylate cyclase. (SW:CYAA_SALTY)" /codon_start=1 /transl_table=11 /product="adenylate cyclase" /protein_id="AAL22784.1" /db_xref="GI:16422504" /translation="MYIYETLQRDLDAINQURVDRALAMGPAFOQVYSLPTLLHY HPLMGLDNGVPSCIFCTPDQTHYNELELYRMTQDPPGEPITGVTYTMG STSVGSCSDLDINVCQSWLDGERQLQKCSLLESWAASGLSEVSYFLIDENR FRNSESGLGDCGCTQHILLDEFTYAVRAGKRLMSWVPCDEEHYDYVNTL YAQVLTPEWLDLGLSLSAEYFGASLWOLYKSIDSPYKAVLTKLLEAYSEYP NPRLAKDIKORLHDEIYVSGLDPCVCMLESTVLTALDPTRLDLVRCFCFLKVC EKLSEACVGRREVLSQVSEWDDARLMDNRANKIDQVREAHNELLDAMMO SYNLRIFARNNLVSASFQDIDGLVTRKLYAFAEPALPKVTLVNPQISPDSEPNLT FIHVPGANRSGWLYNRPANSDSIISHOPEYRNYLNKLVAMFWNGLLSRTHLF IKNGVVDLPKIQEMVADYSHHPLRPAFTPKALYSPCEIRHLAIIVNLEYDPTAAF RNKWHDFDKLDFVFGEEONCLIGSIDLYRNSWNEVTLHFNGEQAMIEALTKIL GQKHQDAAPPDSVEVFCYSOHLRGLIRTRVOOLVSCIELRSLSTQETGRFKLRVS GQWGLFFERLNVSVQKLENAIEFYGAJSHNKLHGLSVQVETNOVKLPVSDGFASEG LIQFFETGDEKFNIIYIDENRRAEVYHCEGSEELVROVSRPQYSSSHDRFTYGS SFINFNLPQFYQIVKTDGRAQVIFPTQFINIVPPANQDHDAPLQOYFS" 2943..3304 /gene="STM3940" 2943..2948 /gene="STM3940" /note="putative RBS for STM3940; RegulonDB:STMSIH003880" 2951..3304 /gene="STM3940" /codon_start=1 /transl_table=11 /product="putative inner membrane protein" /protein_id="AAL22785.1" /db_xref="GI:16422505" /translation="MADAFILIGIVAMVSLGFIILINKLFCFTISAGCLLSLCASNASF QLWDASVWGRWGVKPCGLDVIISCDNYHFLYDLGWELVIGIAFLPPTALMLTCAAILLI NMIMALEYRCAGWRR" 3486..4032 /gene="STM3941" 3486..3491 /gene="STM3941" /note="putative RBS for STM3941; RegulonDB:STMSIH003881" 3490..4032 /gene="STM3941"

```
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL22786.1"
/db_xref="GI:16422506"
/translation="NDVFIKGYMSINNPILVCNKKRQIKALFYSVVAIALMMSILF
LNYSLMKWCVFELILLALAGASAWWSAFSGKQLILNQEGVTLHTRLPYVWHEI
DYVGRSDNTVFLAIFKVDVELYCQRITNEKRNFLSLKHNESNRVNMISLANDLD
YSDSLQDIFKNAVARNLEQ"
4040..4406
/gene="STM3942"
RBS
4040..4045
/gene="STM3942"
CDS
4044..4406
/notes="putative RBS for STM3942; RegulonDB:STMS1H003882"
4044..4406
/gene="STM3942"
/transl_table=11
/product="putative cytoplasmic protein"
/protein_id="AAL22787.1"
/db_xref="GI:16422507"
/translation="WFFVITHHIDVLPATLPTHTCPYCNQDVMILIRQKTRTCG
LSQAKRDKFGIAICNHSNEIKERWSFALQRFTEQKPLNFUFWCRIGFWGWLIS
VMPALFLAWLYLPTFRGW"
complement(4477..4810)
/gene="cyay"
/notes="synonym: STM3943"
complement(4477..4797)
/gene="cyay"
/notes="CYAY protein. (SW:CYAY_SALTY)"
/codon_start=1
/transl_table=11
/product="putative frataxin family transport protein"
/protein_id="AAL22788.1"
/db_xref="GI:16422508"
/translation="WNDSEFHLDALWLTIEERLSDWDCSDIDCEINGVLTLSFE
NGSKLIINRQELPHQWLATQGGYHFDLKGDEWVCDRSGETFWDLLEQAATQAGEK
VSFR"
complement(4805..4810)
/gene="cyay"
/notes="putative RBS for cyay; RegulonDB:STMS1H003883"
complement(4867..5262)
/gene="STM3944"
complement(4867..5256)
/gene="STM3944"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAL22789.1"
/db_xref="GI:16422509"
/translation="WPKMTPINRFAHLIKHLKXALRGVIGSGCIYVIESIMHGTING
VRAAGVLTISAVIGLLSALDIIKAPRAPLILLHLFLCGGCVALMWAMLSALNHFPVN
ASVILITGALFIVILYLPUSFILYRSQR"
complement(5244..5766)
/gene="STM3945"
/notes="pseudogene; 2 frameshifts"
/pseudo
complement(5257..5262)
/gene="STM3944"
/notes="putative RBS for STM3944; RegulonDB:STMS1H003884"
5946..6155
/gene="yifL"
/notes="synonym: STM3946"
5946..5951
/gene="yifL"
/notes="putative RBS for yifL; RegulonDB:STMS1H003886"
5952..6155
/gene="yifL"
/codon_start=1
/transl_table=11
/product="putative outer membrane lipoprotein"
/protein_id="AAL22790.1"
/db_xref="GI:16422510"
```

```
/translation="MKNVFKTLAVLITLPSLTGCGLKGPLYFFPADKNAPPPTKKVDSD
QTQSTMPDKNDRATGDGPSQVNY"
6180..7016
gene
Query Match
Best Local Similarity 62.7%; Score 753; DB 1; Length 21692;
Matches 926; Conservative 0; Mismatches 250; Indels 3; Gaps 2;
1 AGAATAATATGAGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
20235 AGGAACAGGGTGGCTATCGCGCCCCCATTTGAGTGGGCCAGGATATAACGCTTACGCCAC 20176
61 GGACCGGGCTCAAACTCTCTGCTGCCAGAAATGCCGCGAGATCATCAACATATATTAAG 120
20175 GGCCCGGCTCAATTTCTCTGCTGCCAGAACGCCGCTAAGTCTCAACATATATGTTAAA 20115
121 CGATTAACATGCCCGAGATCGGATCGGCTAAACAGGCGACCGGAAACGCTCCCTCCCGCGA 180
20115 TGATCGACATGACCCCGATCGGGTCCGGATAAATCCGCCGCGAGCGCCCTTGACCACGG 20056
181 TGGTCGATGATTAGACATCAAAACCCCAATGGAACAGGTCATAGGCCAGTTCGGCATAT 240
20055 TGGTCGATGATTAGATATCAAAACCGAGATGAAGGTCATAGGCCAGTTCAGCGTAT 19996
241 TTATAGTAGTCTCAATACGCCCGCGCAGATGACTACCAACCCGGTCAATGGTCTGTGG 300
19995 TTATACATAACTCTCAATGCGGCCAGGACAAATGACGATCGTCGATCGTTGGAATCATTT 19936
301 CGAAACCGGACAAAGCGCACCGGATGTCATCCACACCACTAACTCTGCTTCAACGCG 360
19935 CGAAACCGGACAAAGCGCACCGGATGTTACCCACGCGGATAAATCCGCTCTTCCCGC 19876
361 TGACGCCAGAAATCAGTCAGCGGTCCTCATGTTAAAGCAGCAAAACGCGTTTCTCTTGT 420
19875 TGACGCCAGAAATCGGTCAGCGGCCCCCATAGCAAAAGCGCAAAACGCGATTCTCTTGT 19816
421 TCCAGTCTTTTCTGCTGTAACATCGGGTATCTGCCTCTTAAACCACTGTAATTCGT 480
19815 TCCAGTCTTTTCTGCTGTAACATCGGGTCTTTCGGCTCTGTTTACAGGTAAATCGT 19756
481 TTTTCTTTCAGCGTCTGACACACGCTCGCACAGTACGCTATTGTGGCAGCAAAATAGAC 540
19755 TTTTTCGCTCTGCTGACACACAGCGCGCAGTAGCGATTGTGGCAGCAAAATAGAC 19696
541 ACACCGGAGTTCATCATGACCTTAGAATGGTGGTTTGCCTACTGCTGACATCGATCAT 600
19695 AACAGGGAGTTTCGATGACCTTTGAATGGTGGTTTGCCTACTGACCTCAACTCT 19636
601 TTTAACGCTGTCGCCAGGCTCTGTCGAATCAACACTATGACCACTCGCTCAACACCG 660
19635 GCTGAGTCTTTCTCGGGTTCAGCGGCCATCAATACATGACGACGCTATCAACATGG 19576
661 TTTATCCGCGCG--GTGGCGTCTATTGTGGGCTTCAGACCGGACTGGCGATTCAATTTGT 718
19575 ATATCGTGGCGCAGTGGCTTCTATCGCGGACTCCAGACCGGCTGGGGATAGATATCGT 19516
719 GCTGTTTTCGGCTGGGTTGGG-ACGCTATTTTCGGCTCAGTGAATGCGTTTCAAGTGT 777
19515 ACTGTGGGCGTTCGACTGGGTACGCTCTTTTCGGCTCGCTCTCTGCTTTTGAATTC 19456
778 GAAGTGGGCGCGCGGCTTACTTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGG 837
19455 GAAATGGGCTGGCGCGGCTTATCTTATCTGGCTGGGTATCCAGCAATGGCGCGCGCAGG 19396
838 TGCATTTACCTTAATCGCTGGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGC 897
19395 CGCTATCATCTGATATACCTCGCCGACGCAATCGGGGTCGGCTGTTTCAACGCGC 19336
898 AGTTTTGTGAATCTCACCAATCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCCGCA 957
19335 GATATTTGTCAATCTAACCAATCCAAAAGTATTGTGTTTCTTCTGGCGCGCTTTCCGCA 19276
958 ATTCAATCGCGCAACAGCGCGCAACTGATGACAGTATATCGTCTCGGCTGCTACCACTAT 1017
```



```
/note="STMD1.91"
/codon_start=1
/trans1_table=1
/product="92% identity with E. coli acetolactate synthase
ii (ILVg) (SP:P00892) ; contains similarity to Pfam
family PF00205 (Thiamine pyrophosphate enzymes),
score=952.6, E=1.5e-295, N=1"
/protein_id="AAF33483.1"
/db_xref="GI:6960293"
/translation="MNGAQWVHALRAGQVTVFYGGATMPVYDALYDGVVHLLC
RHEQAAMVAYGARSCTGKTCVATSGFATNITGLADLLDSFVVAITQGVAP
FVGFDAFVVDYSLACTSHFVQSLSELEPRIMAEFEVANAGRPGLVDIPKD
IAGSELEPWTFTVANEATFPQADVEQARQMLQAKPKMLYGGVQMAQVAPALRK
FIATQMPVCTCTKLGLGVAEADYPYILGMLGHTKAANFAVOBCDLLIAVGARPDOR
ASLRAHAWYDHPGETIYAPLLAKOLSERKPADSVVTTDVGQHQMSAQHMYTYRPE
VITSGTGCTMGEGFPAAVGAQVARNPTVICSDGSPFNNVOELGTVRKQKLPKI
VLLNQRLGMVQWQOLFQRYSETLITNDPFLMASAFGIPGQHITRQDVEAAL
DMLASBEPGLLHVSIIDELNWNVLVPPGASNSEMLEKLS"
6580..6840
/gene="ilvM"
CDS
6580..6840
/gene="ilvM"
/note="STMD1.90"
/codon_start=1
/trans1_table=1
/product="93% identity with E. coli acetolactate synthase
) ILVg) (SP:P13048)"
/db_xref="AAF33482.1"
/db_xref="GI:6960292"
/translation="MQHGVNSARFNPETLRLVVRHGFQVCSMMWEAATDAQNI
NIELVASPRVDLLFSQLSKLVDVAHVACQSAATSQIRA"
6588..7787
/gene="ilvE"
CDS
6588..7787
/gene="ilvE"
/note="STMD1.89"
/codon_start=1
/trans1_table=1
/product="S. typhimurium branched-chain-amino-acid
transaminase (ILVb) (SP:P15168)"
/protein_id="AAF33481.1"
/db_xref="GI:6960291"
/translation="MTTKADYIWFNGEMVRWEDAKVHVMSHALHYGTSVFEGIRCYD
SHKGPVFRHREHQRLEDSAKIYRFPVSOSIDELMEACHDVIKKNLSAIRPLVF
VGDVGMGNPPGVTVDVILIAFFWGLYALGREALDQIDAVSSWRAAPNTIPTPAK
AGNVLSSLLVGSARRHYQEGIALDVNGYISEGAGENLFEVKGDLVFTPTPTSSAL
PGITDAILKLAELGIEVREQLVRESLVLADEVFMSGTAABITFVRSYDGIQVGE
RCGPVTKRIQAQFLFTGETEDXGWLDPVNS"
7948..9798
/gene="ilvD"
CDS
7948..9798
/gene="ilvD"
/note="STMD1.88"
/codon_start=1
/trans1_table=1
/product="S. typhimurium dihydroxy-acid dehydratase (ILVD)
(SP:P40810) ; contains similarity to Pfam family PF00920
(Dehydratase family), score=1023.9, E=3.5e-304, N=1"
/protein_id="AAF33480.1"
/db_xref="GI:6960290"
/translation="MPKRSATTTTHGRNAGARALWRATGWTDSDFGKPIIAVNSFT
QFVPHVHLRLGLKVAQIEASGVAKENFTIADDDGIAMGHGMLYSLPSRELIIAD
SVYVMNAFCADAMVCISNCKITPGMLASRLNIPIVFSGPMMEAGTKLSKDII
KLDLVDAMIQADPKVSDQSNQOVERSACPTCGSCGMFTANSMNCLTEALGLSPGN
GSLLAHADKQLFLNAGKRIVELTKRYVQNDESLPRNIASKAPFNMTLIDIAMG
GSTNVLHLLAAQEAIEDFTMSIDIKLSRPQLKVPASTOKYHMEVDHVRAGVGLG
ILGELDRAGLNRNVLGUTLPQTLQYDITVQDEAVKRMFRAGPAGIRITQAFS
QDCRWDSLDDRAAGCINSELYASKDGLAVLYGNFAENGCIKTAGVDDSLKFTG
PAKYVESQDDRAALIGCKVVEGDVVIRYEGPKGFMQEMLYPTFSFLKSMGLKAC
ALITDGRFSGGTSLGISHVSPAAAGTIALIDGDTIADIPNASIQLQISEABIA
ARREAQEARGDKAMTPKMRQSVQSFALRAVASLATSADKGAVRDKSLKG"
```

```
Query Match 62.7%; Score 753; DB 1; Length 96086;
Best Local Similarity 78.5%; Pred. No. 4.4e-223;
Matches 926; Conservative 0; Mismatches 250; Indels 3; Gaps 2;

QY 1 AGAATAATGTGGAGATCGCACCCGCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 62264 AGGAACAGGGTGGCTATCGCGCCCCCATTTAGTGGCCAGGATATAACGCTTACGCCAC 62205

QY 61 GGACCGGGCTGAACCTCTGCTGTCAGAAATGCGGCCAGATCATCAACATATCATTAAG 120
Db 62204 GGCCCGGGCTCAATTTCTGCTGTCAGAACCGCGCTAAGTCTCTCAACATATCGTTAAA 62145

QY 121 CGATTAAACATCCCGAGATGCGGATCGGCTTAAACAGGCGACCGGAAAGTCTCCTCCCGCGA 180
Db 62144 TGATCGACATGACCCCGATGCGGGTCCGATTAACATCGGCCGAGCGCTTACCAACGG 62085

QY 181 TGTGCGATGATTAAGACATCAAAACCCCAATGGAAACAGGTATAGCGCCAGTTCGGCATAT 240
Db 62084 TGGTCGATGATGAAGATATCAAAACCGAGATGAAAAAGGTATACATCGCCAGTTCAGCGTAT 62025

QY 241 TTTACGTAGCTCTCAATACGCCCGCGGCAGATGACTACCAACCGGTCATGTCGTCTGCG 300
Db 62024 TTTACATNACTCTCAATGCGGCCAGGACAAATGACGATCGTCCGATCGTTGGATCATTT 61965

QY 301 CGAAAAACGACAAAGCGACCGGAAATCTCATCACACAGTAAATCTCTGCTTCATCACGC 360
Db 61964 CGAAAAACGACAAAGCGACCGGAAATCTTACCCACGCGGATAAATCTCCGCTCTTCCCGC 61905

QY 361 TGAGCCGAGAATCAGTCAGCGGTCCCATGTTAAAGACGAAACGCTTTCTCTTGT 420
Db 61904 TGAGCCGAGAATCGTTCAGCGGCCCATAGCAAAAGCGCAACGATTTCTCTTGT 61845

QY 421 TCCAGTCTTTTCTGCTCTGCTGAAACATCGGTAATCTGCTCTTTAAACACGTAATAATCGT 480
Db 61844 TCCAGTCAATTTCTGCTCTGAAACATCGGTTTTCGCGCTCGTTTACAGGTAAATCGT 61785

QY 481 TTTTATAGCTGCTGACACACGCTGCGACAGTACGCTATTTGTGCGCACAAAAATAGAC 540
Db 61784 TTTTGGCTGCTGCTGACACACAGCGCGAGTAGCGCTATTTGTGCGCACAAAAATAGAC 61725

QY 541 ACACCGGAGTTCATCATGACCTTAGAATGGTGGTTTGGCTACTCTCTGACATCGATCAT 600
Db 61724 AACCGAGGAGTTTCGCTGACCTTTGAAATGGTGGTTTGGCTACTCTGACCTCAACTCT 61665

QY 601 TTTAACGCTGTCGCGAGCTCTGCTGCAATCAACATATGACACCTCTGCTCAACACCG 560
Db 61664 GCTGAGTCTTTCTCCGGGTTTCAGCGCCATCAATACCATGACGACTCTATCAACCATGG 61605

QY 661 TTATCCGCGCG--GTGGCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATTCATATTCT 718
Db 61604 ATATCGTGGCGAGTGGCTTCTATCGCGGACTCCAGACCGGGCTGGGATACATATCGT 61545

QY 719 GCTGGTTGGCGTGGGTTGGG-AGGCTATTTTCCGCTCAGTGATTTGCGTTTGAAGTGT 777
Db 61544 ACTGGTGGCGGCTCGGACTGGGTACGCTCTTTTTCGCGCTCGCTCTGTTTGAATTTCT 61485

QY 778 GAAGTGGCGAGGCGGCTTACTTTGCTGGGTAATCCAGCAGTGGCGCGCGCTGG 837
Db 61484 GAATGGGCTGGCGCGGCTTACTTTTCTGCGTGGGTATCCAGCATGCGCGCGCGCAG 61425

QY 838 TGAATTGAACCTTAATTCGCTGCTCTACTCAATCGCTGACATTTGTTCCAGCGCGC 897
Db 61424 CGCTATCGATCTGCATACCTCTGCCAGACGCAATCCGGGGTTCGGCTGTTTCAAAACGGCG 61365

QY 898 AGTTTTGTGAATCTCAACCAATCCCAAAAGTATTTGTTTCTGGCGCGCTATTTCCGCA 957
Db 61364 GATATTTGTCAATCAACCAATCCCAAAAGTATTTGTTTCTTTCGCGCGCTGTTTCCGCA 61305

QY 958 ATTCAATCGCCCAACAGCCGCAACTGATGCAATATATCGTCTCGCGCTCACCACAT 1017
Db 61304 ATTCATCATCGCCGCAACCGCAACTGCGCGAGTACCTCAITCTCGCGCTCACCACAT 61245

QY 1018 TGTGCTGATATATTGTGATGATCGTTACGCCACCTTGTCTCAACGATGCTCTATG 1077
```


* 32468 32878: contig of 411 bp in length
 * gap of unknown length
 * 32879 34069: contig of 1191 bp in length
 * gap of unknown length
 * 34070 34694: contig of 625 bp in length
 * gap of unknown length
 * 34695 35933: contig of 1239 bp in length
 * gap of unknown length
 * 35934 36920: contig of 987 bp in length
 * gap of unknown length
 * 36921 37173: contig of 253 bp in length
 * gap of unknown length
 * 37174 37282: contig of 109 bp in length
 * gap of unknown length
 * 37283 38196: contig of 914 bp in length
 * gap of unknown length
 * 38197 38452: contig of 256 bp in length
 * gap of unknown length
 * 38453 38678: contig of 226 bp in length
 * gap of unknown length
 * 38679 38835: contig of 157 bp in length
 * gap of unknown length
 * 38836 39561: contig of 726 bp in length
 * gap of unknown length
 * 39562 40452: contig of 891 bp in length
 * gap of unknown length
 * 40453 41223: contig of 771 bp in length
 * gap of unknown length
 * 41224 41864: contig of 641 bp in length
 * gap of unknown length
 * 41865 42690: contig of 826 bp in length
 * gap of unknown length
 * 42691 42988: contig of 298 bp in length
 * gap of unknown length
 * 42989 43612: contig of 624 bp in length
 * gap of unknown length
 * 43613 44265: contig of 653 bp in length
 * gap of unknown length
 * 44266 45263: contig of 998 bp in length
 * gap of unknown length
 * 45264 46592: contig of 1329 bp in length
 * gap of unknown length
 * 46593 47084: contig of 462 bp in length
 * gap of unknown length
 * 47055 47193: contig of 139 bp in length
 * gap of unknown length
 * 47194 48027: contig of 834 bp in length
 * gap of unknown length
 * 48028 48670: contig of 643 bp in length
 * gap of unknown length
 * 48671 49511: contig of 841 bp in length
 * gap of unknown length
 * 49512 49895: contig of 384 bp in length
 * gap of unknown length
 * 49896 50124: contig of 229 bp in length
 * gap of unknown length
 * 50125 50776: contig of 652 bp in length
 * gap of unknown length
 * 50777 50905: contig of 129 bp in length
 * gap of unknown length
 * 50906 51260: contig of 355 bp in length
 * gap of unknown length
 * 51261 52029: contig of 769 bp in length
 * gap of unknown length
 * 52030 52294: contig of 265 bp in length
 * gap of unknown length
 * 52295 53140: contig of 846 bp in length
 * gap of unknown length

Query Match 58.2%; Score 699; DB 2; Length 265383;
 Best Local Similarity 96.3%; Pred. No. 4.2e-206;
 Matches 748; Conservative 0; Mismatches 25; Indels 4; Gaps 3;

QY 200 CAATCCCAATGGAACAGGTCATAGCCAGTTCCGCATATATTTACGTAGCTCTCAATAC 259
 Db 55129 CATAATCCAAATGGAACAGGTCATAGCCAGTT-CGCATATATTTACGTAGCTCTCAATAC 55187
 QY 260 GCCCGGGCAGATGACTACACACCCGGTCATGGTGTGTGCGGAAACCGAACAGCGCA 319
 Db 55188 GCCCGGGCAGATGACTACACACCCGGTCATGGTGTGTGCGGAAACCGAACAGCGCA 55247
 QY 320 CCGGAATGTCTATCCACACACAGTAACTCTGTTTCATCAGCTGACGCCAGAAATCAGTCA 379
 Db 55248 CCGGAATGTCTATCCACACACAGTAACTCTGTTTCATCAGCTGACGCCAGAAATCAGTCA 55307
 QY 380 GCGTCCCATGTTAAAGCAGCAACCGCTTTCTCTTTTCCAGTCTTTTGTGTGCT 439
 Db 55308 GCGTCCCATGTTAAAGCAGCAACCGCTTTCTCTTTTCCAGTCTTTTGTGTGCT 55367
 QY 440 GAAACATCGGTGTAATCTGCTTAAACACAGTAATCGTTTATTTTGTAGCGCTGAC 499
 Db 55368 GAAACATCGGTGTAATCTGCTTAAACACAGTAATCGTTTATTTTGTAGCGCTGAC 55427
 QY 500 ACAACGCTGCGACAGTAGCGTATTGTGGCAAAATAGACACACCGGGAGTTTCATCATG 559
 Db 55428 ACAACGCTGCGACAGTAGCGTATTGTGGCAAAATAGACACACCGGGAGTTTCATCATG 55487
 QY 560 ACCTTAGAAATGGTGGTTTGGCTACCTGTCATCGATCATTTTACGCTGTGCCAGGC 619
 Db 55488 ACCTTAGAAATGGTGGTTTGGCTACCTGTCATCGATCATTTTACGCTGTGCCAGGC 55547
 QY 620 TCTGTGTGCAATCAACACTATGACCACTCTGCTCAACACCGGTTAT--CCGSCCGGTGGCG 677
 Db 55548 TCTGTGTGCAATCAACACTATGACCACTCTGCTCAACACCGGTTATCGCGCGCGGTGGCG 55607
 QY 678 TCTATTGCTGGCTTCAGACCGGACTGGCATTCATATTGTGCTGTGGCGTGGGTT- 736
 Db 55608 TCTATTGCTGGCTTCAGACCGGACTGGCATTCATATTGTGCTGTGGCGTGGGTTG 55667
 QY 737 GGGACGCTATTTCCCGCTCAGTCAATTCGCTTTGAAGTGTGAAGTGGCAGCGCGGCT 796
 Db 55668 GGGACGCTATTTCCCGCTCAGTCAATTCGCTTTGAAGTGTGAAGTGGCAGCGCGGCT 55727
 QY 797 TACTTGATTTGGCTGGGAATCAGCAGTGGGGCGCGCTGGTGCATTTGACCTTAATCG 856
 Db 55728 TACTTGATTTGGCTGGGAATCAGCAGTGGGGCGCGCTGGTGCATTTGACCTTAATCG 55787
 QY 857 CTGGCTCTACTCAATCGCTGCACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 916
 Db 55788 CTGGCTCTACTCAATCGCTGCACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 55847
 QY 917 AATCCAAAAGTATTTGTTCTCGCGCGCTATTTCCGCAATTCATCATGCGCGAA 973
 Db 55848 AATCCAAAAGTATTTGTTCTCGCGCGCTATCGTACCGCGGTGCTCGAA 55904
 RESULT 15
 AEO15396/6
 LOCUS Shigella flexneri 2a str. 301 section 359 of 412 of the complete genome.
 DEFINITION Shigella flexneri 2a str. 301 section 359 of 412 of the complete genome.
 ACCESSION AEO15396.1
 VERSION AEO15396.1 GI:24054379
 KEYWORDS Shigella flexneri 2a str. 301
 SOURCE Shigella flexneri 2a str. 301
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
 REFERENCE 1 (bases 1 to 10253)
 AUTHORS Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H., Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P., Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L., Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
 TITLE Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli


```

gene
5971..6840
/ gene="plda"
/ locus tag="SF3899"
CDS
5971..6840
/ gene="plda"
/ locus tag="SF3899"
/ note="Residues 1 to 289 of 289 are 100 pct identical to
residues 1 to 289 of a 289 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_290453.1"
/ codon_start=1
/ transl_table=11
/ product="outer membrane phospholipase A"
/ protein_id="AAN45334.1"
/ db_xref="GI:2405438"
/ translation="MRTIQGLLPLVFMPLMAVYAEATVKEVHDAVAVGSIANNLQ
EHDNFTLYPYDNYLYTQTSLDNKIAIASYDAENARKDEVKFLSLAFPLMRGIL
GPNVLGASYQKSWLQSNSEESPPFRETNYEQLFLGPDYRFAGWTLRDVENGY
NHDNSGRDPTSRNWRLYTLMAENGLVEKPVYVGNVDNDNDPIIKMYGYOLK
IGYHIGDAVLAKQYNNWNTYGGAEGLSVPIKTHVRLYTVYVSGVGSLLIDYNFQ
TRVGVGVMNDLF"
5988..8802
/ gene="recQ"
/ locus tag="SF3900"
5988..8802
/ gene="recQ"
/ locus tag="SF3900"
/ note="Residues 1 to 604 of 604 are 99 pct identical to
residues 8 to 611 of a 611 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_290454.1"
/ codon_start=1
/ transl_table=11
/ product="ATP-dependent DNA helicase"
/ protein_id="AAN45335.1"
/ db_xref="GI:2405438"
/ translation="MLNLESKAKVQLQETPGYQDFRGQBEIITVLSGRDCLVYMP
GGKSLCTQIFALLNGLVTVVPSLSMDQVQLQANGVAACLNSTQREQLEV"

```

Search completed: March 28, 2004, 16:44:55
Job time : 5039.17 secs

```

Query Match 57.5%; Score 690; DB 1; Length 10253;
Best Local Similarity 98.2%; Pred. No. 1.4e-203;
Matches 719; Conservative 0; Mismatches 10; Indels 3; Gaps 2;

QY 472 TAAATCGTTTTTTTATAGCGCTGACACACCGCTGACAGTAGCGTATTTGGCACA 531
DB 10253 TAAATCGTTTTTTTAAAGCGTCCCTGACACACCGCTGACAGTAGCGTATTTGGCACA 10194

QY 532 AAAATAGACACACCGGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCTTACCTGCTGAC 591
DB 10193 AAAATAGACACACCGGGAGTTTCATCATGACCTTAGAATGGTGGTTTGCTTACCTGCTGAC 10134

QY 592 ATCGATCATTTTAAGCTGTGCGCAGGCTCTGGTGCAATCAACACTATGACCACTCGCT 651
DB 10133 ATCGATCATTTTAAGCTGTGCGCAGGCTCTGGTGCAATCAACACTATGACCACTCGCT 10074

QY 652 CAACACCGGTAT--CCGGCGGTGGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGAT 709
DB 10073 CAACACCGGTATTCGGCGCGGTGGGCTCTATTGCTGGGCTTCAGACCGGACTGGCGAT 10014

QY 710 TCATATTGCTGGTTGGGCTGGGTT--GGAGCGTATTTTCCCGCTCAGTGATTCGCTT 768
DB 10013 TCATATTGCTGGTTGGGCTGGGTTGGGACCGCTATTTTCCCGCTCAGTGATTCGCTT 9954

QY 769 TGAAGTGTGAGTGGGCGAGCGGCTTACTGTATTTGGCTGGGAATCCAGCAGTGGCG 828
DB 9953 TGAAGTGTGAGTGGGCGAGCGGCTTACTGTATTTGGCTGGGAATCCAGCAGTGGCG 9894

QY 829 CGCGCTGTGCAATTGACCTTAAATCGCTGCTCTACTCAATCGGCTGCAATTTGTT 888
DB 9893 CGCGCTGTGCGCAATTGATCTTAAATCGCTGCTCTACTCAATCGGCTGCAATTTGTT 9834

QY 889 CCAGCGCAGTTTGTGAATCTCACCACCAATCCCAAGATATGTGTTTCTGGCGGCT 948
DB 9833 CCAGCGCAGTTTGTGAATCTCACCACCAATCCCAAGATATGTGTTTCTGGCGGCT 9774

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 12:14:43 ; Search time 597.686 Seconds
(without alignments)
8529.292 Million cell updates/sec

Title: US-09-847-392-1

Perfect score: 1200

Sequence: 1 agaaataatgtgagatcgc.....atctcgagatcgccgtaaac 1200

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1200	100.0	1200	3	Aa294405 E. coli r
2	1200	100.0	1231	3	Aa48442 E. coli L
3	571.2	47.6	622	5	Aa893395 DNA encod
4	523.6	43.6	2216	5	Aa889415 DNA encod
5	446	37.2	1023	7	Aa332708 Prokaryot
6	417	34.8	417	4	Aa84646 E. coli g
7	417	34.8	417	7	Aa18626 Prokaryot
8	382.2	31.9	621	7	Aa51728 Prokaryot
9	375	31.2	618	7	Aa36103 Prokaryot
10	357.4	29.8	618	7	Aa32157 Prokaryot
11	348.2	29.0	623	7	Aa448809 Prokaryot
12	298.8	24.9	1017	7	Aa19693 Prokaryot
13	297.2	24.8	1017	7	Aa51727 Prokaryot
14	291.4	24.3	2054	5	Aa88021 DNA encod
15	290.8	24.2	990	7	Aa32154 Prokaryot
16	276.4	23.0	990	7	Aa35868 Prokaryot
17	275	22.9	621	7	Aa53970 Prokaryot
18	262.8	21.9	1026	7	Aa448808 Prokaryot
19	259	21.6	359	7	Aa13399 Prokaryot
20	254.2	21.2	4303	5	Aa85519 DNA encod
21	248.6	20.7	1884	5	Aa889412 DNA encod
22	243	20.2	1032	5	Aa885510 DNA encod
23	243	20.2	1746	5	Aa874968 DNA encod

C 24	243	20.2	2410	5	Aa888022	Aa88022 DNA encod
C 25	207.4	17.3	283	4	Aa884479	Aa884479 E. coli g
C 26	206	17.2	2410	5	Aa888022	Aa88022 DNA encod
C 27	205.2	17.1	1032	5	Aa885510	Aa885510 DNA encod
C 28	181.8	15.2	1014	7	Aa53968	Aa53968 Prokaryot
C 29	175	14.6	274	5	Aa89410	Aa89410 DNA encod
C 30	167.8	14.0	516	5	Aa893398	Aa893398 DNA encod
C 31	149.8	12.5	618	7	Aa52910	Aa52910 Prokaryot
C 32	144.8	12.1	1002	7	Aa44944	Aa44944 Prokaryot
C 33	141	11.8	1041	5	Aa893397	Aa893397 DNA encod
C 34	133	11.1	110000	7	ACF67367_37	Continuation (38 o
C 35	133	11.1	110000	7	ACF67367_38	Continuation (39 o
C 36	133	11.1	110000	7	ACF65388_10	Continuation (38 o
C 37	132.4	11.0	999	7	ACF70714	Continuation (38 o
C 38	112.4	9.4	630	7	ACA42760	Ac42760 Prokaryot
C 39	104.2	8.7	1746	5	Aa874968	Aa874968 DNA encod
C 40	104	8.7	630	7	ACA45284	ACA45284 Prokaryot
C 41	100	8.3	100	7	ACD71734	ACD71734 E. coli K
C 42	100	8.3	100	7	ACD71736	ACD71736 E. coli K
C 43	100	8.3	100	7	ACD71735	ACD71735 E. coli K
C 44	86.8	7.2	633	7	ACA43829	ACA43829 Prokaryot
C 45	80	6.7	627	7	ACA27048	ACA27048 Prokaryot

ALIGNMENTS

RESULT 1

AZ94405
ID AAZ94405 standard; DNA; 1200 BP.

XX

AC AAZ94405;

XX

DT 18-JUL-2000 (first entry)

XX

DE E. coli rhtB gene imparting homoserine resistance.

XX

KW Homoserine resistance; rhtB gene; L-homoserine; L-alanine; L-isoleucine;

KW L-valine; L-threonine; ss.

XX

OS Escherichia coli.

XX

PH Key Location/Qualifiers

FT CDS 557..1171

FT /*tag= a

FT /note= "a DNA sequence corresponding to the coding region (minus the stop codon) is specifically claimed in Claim 3"

XX

XX EP994190-A2.

XX

XX 19-APR-2000.

XX

XX 20-SEP-1999; 99EP-00118591.

XX

XX 13-OCT-1998; 98RU-00118425.

XX

XX (AJIN) AJINOMOTO CO INC.

XX

XX Livshits VA, Zakataeva NP, Aleoshin VV, Belareova AV;

XX Tokhmakova IL;

XX

XX WPI; 2000-273530/24.

XX

XX P-PSDB; AAY79298.

XX

XX Novel RhtB protein, useful for generation of L-homoserine resistance in Escherichia bacteria and large-scale production of e.g. L-homoserine and L-alanine.

XX

XX Claim 3; Page 10-11; 14pp; English.

XX

XX This is the DNA sequence the novel rhtB gene of Escherichia coli K-12.

XX The rhtB Gene maps at 86 min on the E. coli chromosome. The gene

CC	participates in resistance to homoserine. Amplification of the gene
CC	results in an improvement of the amino acid productivity of <i>E. coli</i> . The
CC	invention provides: the RhtB protein (see AA7998); DNA encoding the
CC	RhtB, especially nucleotides 557-1171 of the present sequence; a
CC	bacterium, especially of the genus <i>Escherichia</i> , in which L-homoserine
CC	resistance is enhanced by amplifying the copy number or increasing the
CC	expression rate of the <i>rhtB</i> DNA, the DNA being carried on a multicopy
CC	vector or on a transposon; and a method for producing an amino acid by
CC	cultivating the bacterium in a culture medium to produce and accumulate
CC	the amino acid in the medium, from which it is recovered. The method is
CC	used for the production of L-homoserine, L-alanine, L-isoleucine, L-
CC	valine or L-threonine (all claimed)
XX	
SQ	Sequence 1200 BP; 285 A; 309 C; 305 G; 301 T; 0 U; 0 Other;
	Query Match 100.0%; Score 1200; DB 3; Length 1200;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 AGAATAATGTGGAGATCGACGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 60
Db	1 AGAATAATGTGGAGATCGACGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 60
Qy	61 GGACCGGGCTGAACCTCCTGTGCCAATGCGGCCAGATCATCAACATAATCATTTAAAG 120
Db	61 GGACCGGGCTGAACCTCCTGTGCCAATGCGGCCAGATCATCAACATAATCATTTAAAG 120
Qy	121 CGATTAAATGCGGAGATCGGATCGGCTAAACGGCGACGGAAAGTCCTGCCCGCGA 180
Db	121 CGATTAAATGCGGAGATCGGATCGGCTAAACGGCGACGGAAAGTCCTGCCCGCGA 180
Qy	181 TGGTCGATGATTAAAGACATCAACCCCAATGGAACAGGTCATAGGCCAGTTCGGCATAT 240
Db	181 TGGTCGATGATTAAAGACATCAACCCCAATGGAACAGGTCATAGGCCAGTTCGGCATAT 240
Qy	241 TTTACGTAGCTCTCAATAGCCCCGGGAGATGACTACCAACCGGTCATGGTCTGTGG 300
Db	241 TTTACGTAGCTCTCAATAGCCCCGGGAGATGACTACCAACCGGTCATGGTCTGTGG 300
Qy	301 CGAAAACGACAAACGGCACCGGAATGTCTACACACAGGTAAACTCTGCTTCATCAACGC 360
Db	301 CGAAAACGACAAACGGCACCGGAATGTCTACACACAGGTAAACTCTGCTTCATCAACGC 360
Qy	361 TGACGCCAGAAATCAGTCAGCGGTCCCATGTGTAAAGACGAAACGGTTTTCTTTGTT 420
Db	361 TGACGCCAGAAATCAGTCAGCGGTCCCATGTGTAAAGACGAAACGGTTTTCTTTGTT 420
Qy	421 TCCAGTCTTTTCTGCTGTAACATCGGTAATCTGCCTCTTAAACACAGTAAATCGT 480
Db	421 TCCAGTCTTTTCTGCTGTAACATCGGTAATCTGCCTCTTAAACACAGTAAATCGT 480
Qy	481 TTTTTTTAGCGTGCCTGACAAACGCTGCGACAGTAGCGTATTGTGGCAAAATAAGAC 540
Db	481 TTTTTTTAGCGTGCCTGACAAACGCTGCGACAGTAGCGTATTGTGGCAAAATAAGAC 540
Qy	541 ACACCGGAGTTCATCATGACCTTAGATGGTGTGTTGCTACCTGTGACATCGATCAT 600
Db	541 ACACCGGAGTTCATCATGACCTTAGATGGTGTGTTGCTACCTGTGACATCGATCAT 600
Qy	601 TTTAAACGCTGTCCGAGGCTCTGGTGCAATCAACACTATGACCAACCTCGTCAACACCG 660
Db	601 TTTAAACGCTGTCCGAGGCTCTGGTGCAATCAACACTATGACCAACCTCGTCAACACCG 660
Qy	661 TTATCCGCGGTGCGCTCTATTGCTGGGCTTCAGACCGGATCGGATCATATTTGTC 720
Db	661 TTATCCGCGGTGCGCTCTATTGCTGGGCTTCAGACCGGATCGGATCATATTTGTC 720
Qy	721 TGGTTGGGCGTGGGGTTGGGAACGCTATTTTCCCGCTCAGTGATTGCGTTGAAGTGA 780
Db	721 TGGTTGGGCGTGGGGTTGGGAACGCTATTTTCCCGCTCAGTGATTGCGTTGAAGTGA 780
Qy	781 GTGGGACGCGCGCTTACTTGATTGGCTGGGAATCCAGCAAGTGGCGCGCTGTGTC 840

Db	781	GTGGCGAGCGCGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC	840		
Qy	841	AATTGACCTTAAATCGCTGGCTCTACTCAATCGGTCGA	900		
Db	841	AATTGACCTTAAATCGCTGGCTCTACTCAATCGGTCGA	900		
Qy	901	TTTTTGGAATCTCACCACATCCCAAAGTATTGTGTTCTGGCGCGCTATTTCGGCAATT	960		
Db	901	TTTTTGGAATCTCACCACATCCCAAAGTATTGTGTTCTGGCGCGCTATTTCGGCAATT	960		
Qy	961	CATCATSCCGCAACGCGCAACTGATGCAGTATATCGTCTCGGCGTCAACACTATTGT	1020		
Db	961	CATCATSCCGCAACGCGCAACTGATGCAGTATATCGTCTCGGCGTCAACACTATTGT	1020		
Qy	1021	GGTCGATATTATTGTCATGATCGGTTACGCCACCTTCTCAACGCGATTGCTCTATGGAT	1080		
Db	1021	GGTCGATATTATTGTCATGATCGGTTACGCCACCTTCTCAACGCGATTGCTCTATGGAT	1080		
Qy	1081	TAAAGGACCAAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT	1140		
Db	1081	TAAAGGACCAAAAGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGT	1140		
Qy	1141	GGAGGCGCTGTTAGCATCGCGGAGCGCTGCGTGAATAATGTCGGATCGGCGTAAAC	1200		
Db	1141	GGAGGCGCTGTTAGCATCGCGGAGCGCTGCGTGAATAATGTCGGATCGGCGTAAAC	1200		
RESULT 2					
AA48442					
ID	AA48442 standard; DNA; 1231 BP.				
XX					
AC	AA48442;				
XX					
DT	08-SEP-2000 (first entry)				
XX					
DE	E. coli L-homoserine resistance gene, rhtB.				
XX					
KW	L-homoserine resistance; L-homoserine synthesis; rhtB; L-threonine;				
KW	L-valine; L-leucine; ds.				
XX					
OS	Escherichia coli.				
XX					
PH	Key	Location/Qualifiers			
FT	CDS	557..1174			
FT		/*tag= a			
FT		/product= "rhtB"			
XX					
PN	EP1013765-A1.				
XX					
PD	28-JUN-2000.				
XX					
PF	20-DEC-1999; 99EP-00125406.				
XX					
PR	23-DEC-1998; 98RU-00129511.				
XX					
PA	(AJIN) AJINOMOTO KK.				
XX					
PI	Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;				
XX					
DR	WPI; 2000-414602/36.				
DR	P-PSDB; AAY99597.				
XX					
PT	Novel Escherichia bacterium having enhanced L-threonine resistance due to				
PT	enhanced RhtC protein activity, used to produce L-threonine, L-				
PT	homoserine, L-valine and L-leucine.				
XX					
PS	Claim 4; Page 12-13; 24pp; English.				
XX					
CC	The present sequence is the L-homoserine resistance gene, rhtB, from				
CC	Escherichia coli. This sequence may be used to impart L-homoserine				
CC	resistance on E. coli bacteria, which would be useful for producing a				
CC	high yield of L-homoserine. L-homoserine resistance means that the				
CC	bacteria will be able to grow on a minimal medium containing L-homoserine				

CC at a concentration at which the corresponding wild-type strain would not
CC grow. Since the transformed bacteria can grow on the minimal medium, it
CC can synthesize L-homoserine, which accumulates. The accumulated amino
CC acids can then be removed from the culture medium. The bacterium of the
CC present invention may also be used to synthesize L-threonine, L-valine
CC and L-leucine at increased levels

XX
SQ Sequence 1231 BP; 291 A; 319 C; 311 G; 310 T; 0 U; 0 Other;

Query Match 100.0%; Score 1200; DB 3; Length 1231;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATATATGTCGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60

DB 1 AGAAATATATGTCGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60

QY 61 GGACCGGCTGAACCTCTGTCGCCAGATCGCCAGATCATCAACATATCATTAAG 120

DB 61 GGACCGGCTGAACCTCTGTCGCCAGATCGCCAGATCATCAACATATCATTAAG 120

QY 121 CGATTAAATGCGGAGATCGGATCGGCTAACAGGCGACCGGAACCTCCCTGCCCGCGA 180

DB 121 CGATTAAATGCGGAGATCGGATCGGCTAACAGGCGACCGGAACCTCCCTGCCCGCGA 180

QY 181 TGGTCGATGATTAAACATCAACCCCAATGGAACAGGTATAGGCGAGTTCGGCATAT 240

DB 181 TGGTCGATGATTAAACATCAACCCCAATGGAACAGGTATAGGCGAGTTCGGCATAT 240

QY 241 TTTACGTAGCTCTCAATACGCCCCGGGCGAGTACTACACCCGGTCATGCTGTGG 300

DB 241 TTTACGTAGCTCTCAATACGCCCCGGGCGAGTACTACACCCGGTCATGCTGTGG 300

QY 301 CGAAACGCGAAGCGACCGGATGTCACACACGATCACTGCTGCTTATCATCGC 360

DB 301 CGAAACGCGAAGCGACCGGATGTCACACACGATCACTGCTGCTTATCATCGC 360

QY 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGTTAAAGCAGCAACGGGTTTCTTTGTT 420

DB 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGTTAAAGCAGCAACGGGTTTCTTTGTT 420

QY 421 TCCAGTCTTTTGTGCTGAAACATCGGTAATCTGCTTAAACACGTAATATCGT 480

DB 421 TCCAGTCTTTTGTGCTGAAACATCGGTAATCTGCTTAAACACGTAATATCGT 480

QY 481 TTTTITTAGCTGCTGACACACGCTCGACAGTAGGTTATGTCGACAAAAATAGAC 540

DB 481 TTTTITTAGCTGCTGACACACGCTCGACAGTAGGTTATGTCGACAAAAATAGAC 540

QY 541 ACACCGGAGTTTCATCATGACCTTAGAATGTGTTTGCCTTACCTGTGATCAT 600

DB 541 ACACCGGAGTTTCATCATGACCTTAGAATGTGTTTGCCTTACCTGTGATCAT 600

QY 601 TTTACGCTGTCGCGAGCTCTGGTGCATCAACACTATGACACCTCGCTCAACACGG 660

DB 601 TTTACGCTGTCGCGAGCTCTGGTGCATCAACACTATGACACCTCGCTCAACACGG 660

QY 661 TTATCCGCGGCTGCTGCTATGTCGGGCTTCAGACCGGACTGGCGATTCAATATGTC 720

DB 661 TTATCCGCGGCTGCTGCTATGTCGGGCTTCAGACCGGACTGGCGATTCAATATGTC 720

QY 721 TGGTGGGCTGGGTTGGACGCTATTTCCGCTCAGTATGCTGTTGAAGTGTAA 780

DB 721 TGGTGGGCTGGGTTGGACGCTATTTCCGCTCAGTATGCTGTTGAAGTGTAA 780

QY 781 GTGGGCGAGCGCGGCTTACTTGAATTTGGCTGGGAATCAGCAGTGGCGCGCTGTGTC 840

DB 781 GTGGGCGAGCGCGGCTTACTTGAATTTGGCTGGGAATCAGCAGTGGCGCGCTGTGTC 840

QY 841 AATTGACCTTAATCGTGGCTCTACTCAATCGGCTGCATTTGTTCCAGCGCGAGT 900

DB 841 AATTGACCTTAATCGTGGCTCTACTCAATCGGCTGCATTTGTTCCAGCGCGAGT 900

QY 901 TTTTGTGAATCTCACCATCCCAAAAGTATTGTGTTCTGGGCGGCTATTTCGCAATT 960

DB 901 TTTTGTGAATCTCACCATCCCAAAAGTATTGTGTTCTGGGCGGCTATTTCGCAATT 960

QY 961 CATCATCGCGCAACAGCGCAACTGATGAGTATATCGTCTGGGCTCACCACTATTGT 1020

DB 961 CATCATCGCGCAACAGCGCAACTGATGAGTATATCGTCTGGGCTCACCACTATTGT 1020

QY 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCTTGTCAACGGATTGCTTATGGAT 1080

DB 1021 GGTGATATTATTGTGATGATCGGTTACGCCACCTTGTCAACGGATTGCTTATGGAT 1080

QY 1081 TAAAGGCCAAGCAGATGAAGCGCTGAATATGTTTCGGCTGTTTATGCTGTT 1140

DB 1081 TAAAGGCCAAGCAGATGAAGCGCTGAATATGTTTCGGCTGTTTATGCTGTT 1140

QY 1141 GGGAGCGCTGTAGCATCGCGGAGGATGCGTGAATAAATGCGATGCGGCGTAAAC 1200

DB 1141 GGGAGCGCTGTAGCATCGCGGAGGATGCGTGAATAAATGCGATGCGGCGTAAAC 1200

RESULT 3

AAS93395

ID AAS93395 standard; cDNA; 622 BP.

XX AC AAS93395;

XX AC AAS93395;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #29199.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG29208.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity.

XX Claim 1; SEQ ID NO 29199; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridization probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 622 BP; 123 A; 152 C; 172 G; 175 T; 0 U; 0 Other;

Query Match 47.6%; Score 571.2; DB 5; Length 622;

Best Local Similarity 98.1%; Pred. No. 8.7e-179;

Matches 610; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

QY 557 ATGACCTTAGATGCTGGTTCCTACCTGCTGATCGATCATATTTAA-CGCTGTGCGC 615

Db 1 ATGACCTTAGAATGGTGTTCCTACCTGCTGATCGATCATATTTAAAGCCGTTCGCC 60

QY 616 AGGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTAT--CGGCGCGT 673

Db 61 AGGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTATCGCGCGGT 120

QY 674 GCGGTCTATTGCTGGCTTCAGACCGGACTGGGATTCATATTTGCTGTTGGCGTGG 733

Db 121 GCGGTCTATTGCTGGCTTCAGACCGGACTGGGATTCATATTTGCTGTTGGCGTGG 180

QY 734 GTT-GGGAGCTATTTCCTGCTAGTTCGTTTGAAGTGTGAAGTGGCAGGCGC 792

Db 181 GTTGGGAGCTATTTCCTGCTAGTTCGTTTGAAGTGTGAAGTGGCAGGCGC 240

QY 793 GGTCTACTGATTGGCTGGGAATCCAGCATGCGCGCGCTGGTGCAATGACCTTAA 852

Db 241 GGTCTACTGATTGGCTGGGAATCCAGCATGCGCGCGCTGGTGCAATGACCTTAA 300

QY 853 ATCGCTGGGCTCTACTCAATCGCTGACATTTCTTCAGCGCGCAGTTTGTGAATCT 912

Db 301 ATCGCTGGGCTCTACTCAATCGCTGACATTTCTTCAGCGCGCAGTTTGTGAATCT 360

QY 913 CACCAATCCAAAGATTTGTTTCTGGCGCGCTATTTCCGCAATTCATGCGCA 972

Db 361 CACCAATCCAAAGATTTGTTTCTGGCGCGCTATTTCCGCAATTCATGCGCA 420

QY 973 ACAGCGCACTGATGAGTATCGTCTCGGCGTCAACACTATTTGTCGATATTAT 1032

Db 421 ACAGCGCACTGATGAGTATCGTCTCGGCGTCAACACTATTTGTCGATATTAT 480

QY 1033 TGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGATTAAGGACCAA 1092

Db 481 TGTGATGATCGGTTACGCCACCCCTTGCTCAACGGATTGCTCTATGATTAAGGACCAA 540

QY 1093 GCAGATGAAGCGCTGAATAGATTTTCGGCTCGTTGTTATGCTGTGGAGCGCTGTT 1152

Db 541 GCAGATGAAGCGCTGAATAGATTTTCGGCTCGTTGTTATGCTGTGGAGCGCTGTT 600

QY 1153 AGCATCGGCGAGCGCATGGTGA 1174

Db 601 AGCATCGGCGAGCATGGTGA 622

RESULT 4

ID AAS89415/c

XX AAS89415 standard; cDNA; 2216 BP.

XX AAS89415;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #25219.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PsDB; ABG25228.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 25219; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2216 BP; 518 A; 562 C; 639 G; 497 T; 0 U; 0 Other;

Query Match 43.6%; Score 523.6; DB 5; Length 2216;

Best Local Similarity 91.3%; Pred. No. 1.3e-162;

Matches 623; Conservative 0; Mismatches 49; Indels 10; Gaps 6;

QY 527 GCACAAAATAGACACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTCCTACCTG 586

Db 2116 GCACAAAATAGACAC-CCCGGAGTTTCATCATG-CCTTAGAAAAGGGGTTGCCACCCG 2059

QY 587 CTGACATCGATCATTTTAAAGCTGTGCGCAGGCTCTGGTCAATCAACACTACACACC 646

Db 2058 GACATGATCTTTTATAGCCCCCTTCCCAGTTTCTGGCCCAATCACCATTATGCCCCC 1999

QY 647 TCGCTCAACACCGGTTAT--CCGCGCGGTGGCGTCTATTCTGGGCTTCAGACCGGACTG 704

Db 1998 TCGCTCAACACCGGTTATCGCGCGGTGGCGTCTATTCTGGGCTTCAGACCGGACTG 1939

QY 705 GCGATTCTATTTGCTGGTGGCGTGGGTT--GGGACGCTATTTTCCCGCTCAGTGATT 763

Db 1938 GCGATTCTATTTGCTGGTGGCGTGGGTTGGGACGCTATTTTCCCGCTCAGTGATT 1879

QY 764 GCGTTTGAAGTGTGAAGTGGGAGCGCGGCTTACTTGTGGCTGGGAATCCAGCAG 823

Db 1878 GCGTTTGAAGTGTGAAGTGGGAGCGCGGCTTACTTGTGGCTGGGAATCCAGCAG 1819

QY 824 TGGCGCGCGCTGGTGCAATTGACCTTAATTCGCTGGCTCTACTCAATCGCGTGCACAT 883

Db 1818 TGGCGCGCGTGGTGGCAATTAACCTTAATCGTGGCTTACTCAATCGCGTCGACAT 1759
QY 884 TTGTTTCCAGCG----CGCAGTTTGTGAATCTCAACAATCCCAAAAGTATTGTGTTTCT 939
Db 1758 TTGTTCTACAGCGGCACATTTTGTGAATCTCTCAATCCCAAAAGTATTGTGTTTCT 1699
QY 940 GCGCGCGTATTTCGGAATCAATCGCGCAACAGCGCAACTGATGAGTATATCGT 999
Db 1698 GCGCGCGTATTTCGGAATCAATCGCGCAACAGCGCAACTGATGAGTATATCGT 1639
QY 1000 -GCTCGCGGTACCACTATTGTGTCGATATTATTGTGATGATCGGTACGCCACCTTG 1058
Db 1638 GCGTGGCGTCACTATTGTGTCGATATTATTGTGATGATCGGTACGCCACCTTG 1579
QY 1059 CTCACGGAATGCTCTATGGAATTAAGCAACCAAGCGAGTGAAGCGCTGAATAAGATT 1118
Db 1578 CTCACGGAATGCTCTATGGAATTAAGCAACCAAGCGAGTGAAGCGCTGAATAAGATT 1519
QY 1119 TCGGCTCGTTGTTATGCTGTGGGCGCTGTAGCATCGCGAGCGATCGTGAATAA 1178
Db 1518 TCGGCTCGTTGTTATGCTGTGGGCGCTGTAGCATCGCGAGCGATCGTGAATAA 1459
QY 1179 TAATCGGATGCGCGTAAC 1200
Db 1458 TAATCGGATGCGCGTAAC 1437

RESULT 5
AC32708/c
ID ACA32708 standard; DNA; 1023 BP.
AC ACA32708;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #14365.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Escherichia coli.
XX
XX WO200277183-A2.
XX
PD 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
PF 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen XL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX P-PSDB; ABU28838.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 20578; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 623 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIFO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1023 BP; 219 A; 261 C; 289 G; 254 T; 0 U; 0 Other;
XX
Query Match 37.2%; Score 446; DB 7; Length 1023;
Best Local Similarity 100.0%; Pred. No. 5,1e-137;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 60
Db 446 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 387
QY 61 GGACCGGCGTGAACCTCTCTGCTCCAGAAATGCCGCCAGATCATCAACATAATCAATTAAG 120
Db 386 GGACCGGCGTGAACCTCTCTGCTCCAGAAATGCCGCCAGATCATCAACATAATCAATTAAG 327
QY 121 CGATTAAACATGCCCGAGATCGGATCGGCTAACAGCGACCGAAGCGTCCCTGCCGCGA 180
Db 326 CGATTAAACATGCCCGAGATCGGATCGGCTAACAGCGACCGAAGCGTCCCTGCCGCGA 267
QY 181 TGGTCGATGATTAGACATCAACACCCCAATGGAACAGGTCTAGGCCAGTTCCGCATAT 240
Db 266 TGGTCGATGATTAGACATCAACACCCCAATGGAACAGGTCTAGGCCAGTTCCGCATAT 207
QY 241 TTTACGTAGTCTCAATAGCCTCCCGGCGAGATGACTACACCCCGGTCTGCTGTGCG 300
Db 206 TTTACGTAGTCTCAATAGCCTCCCGGCGAGATGACTACACCCCGGTCTGCTGTGCG 147
QY 301 CGAAACGGCAAAAGCGCACCGGAATGTCTCCACACAGTAAACTCTGCTTCATCAAGC 360
Db 146 CGAAACGGCAAAAGCGCACCGGAATGTCTCCACACAGTAAACTCTGCTTCATCAAGC 87
QY 361 TGACCGCAGAAATCAGTCAGCGGTCCCATGTGTAAGAGCAGCAAGCGGTTTCTCTTGT 420
Db 86 TGACCGCAGAAATCAGTCAGCGGTCCCATGTGTAAGAGCAGCAAGCGGTTTCTCTTGT 27
QY 421 TCCAGTCTTTTGTGCTGTAACAT 446
Db 26 TCCAGTCTTTTGTGCTGTAACAT 1
RESULT 6
AAH84645
ID AAH84646 standard; DNA; 417 BP.
XX
AC AAH84646;
XX

26-SEP-2001 (first entry)
E. coli growth and proliferation related coding sequence SEQ ID NO:274.
Escherichia coli; growth; proliferation; microbial; antimicrobial;
bacterial infection; microorganism; ds.
Escherichia coli.
WO200134810-A2.
17-MAY-2001.
09-NOV-2000; 2000WO-US030950.
09-NOV-1999; 99US-0164415P.
(ELIT-) ELITRA PHARM INC.
Forsyth RA, Ohlsen K, Zyskind J;
WPI; 2001-335933/35.
P-PSDB; AAG98975.
Novel nucleic acids that inhibit Escherichia coli proliferation, useful
for screening for homologous genes and for designing expression vectors.
Claim 9; Page 351; 522pp; English.
AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli growth
and proliferation related proteins given in AAG99078 and AAG98830 to
AAG98999. (1) can be used as potential targets for the generation of new
antimicrobial agents, and for identification of compounds which interact
with the gene products of (1). In addition the expression of (1) and the
purification of the proteins, the purified proteins can be used to
generate reagents and screen small molecule libraries or other candidate
compound libraries for compounds that can be further developed to yield
novel antimicrobial compounds. In addition, nucleic acid probes
complementary to (1) that are specific for particular species of
microorganisms can be used to identify particular microorganism species
in clinical specimens, therefore, providing a rapid and dependable method
by which to identify the causative agents of a bacterial infection. Also,
antibodies generated against proteins translated from mRNA transcribed
from proliferation-required sequences can also be used to screen for
specific microorganisms that produce such proteins in a species-specific
manner. AAH84371 and AAH84670 represent sequencing primers used in the
isolation of E. coli growth and proliferation related sequence, which are
used in an example from the present invention
Sequence 417 BP; 87 A; 96 C; 116 G; 118 T; 0 U; 0 Other;

Query Match 34.8%; Score 417; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.3e-127;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
758 GTGATTGCGTTGAAGTTGTAAGTGGCGAGCGCGCTTACTTGTATTTGGCTGGGAATC 817
1 GTGATTGCGTTGAAGTTGTAAGTGGCGAGCGCGCTTACTTGTATTTGGCTGGGAATC 60
818 GAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATCGTGGCGCTTACTCAATGCGGT 877
61 GAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAATCGTGGCGCTTACTCAATGCGGT 120
878 CGACATTGTTCCAGCGCGAGTTTGTGTAATCTCAACATCCAAAGTAATGTGTTT 937
121 CGACATTGTTCCAGCGCGAGTTTGTGTAATCTCAACATCCAAAGTAATGTGTTT 180
938 CTGGCGGCGCTATTTCCGAATTCATATGCGGAGCGGCACTGATGAGTATATC 997
181 CTGGCGGCGCTATTTCCGAATTCATATGCGGAGCGGCACTGATGAGTATATC 240
998 GTGCTGGCGGTCAACCACTATTGTCGATATTTGATGATCGGTACGCCACCCCT 1057

Db 241 GTGCTGGCGGTCAACCACTATTGTCGATATTTGATGATCGGTACGCCACCCCT 300
Qy 1058 GCTCAACGGATTGCTCTATGATTTAAAGGACCAAGCAGATGAAGCGCTGAATAGATT 1117
Db 301 GCTCAACGGATTGCTCTATGATTTAAAGGACCAAGCAGATGAAGCGCTGAATAGATT 350
Qy 1118 TTGGCGCTGTTGTTTATGCTGTTGGGAGCGCTTTAGCATGCGGAGGCGATGCGTGA 1174
Db 361 TTGGCGCTGTTGTTTATGCTGTTGGGAGCGCTTTAGCATGCGGAGGCGATGCGTGA 417
RESULT 7
ACA18626
ID ACA18626 standard; DNA; 417 BP.
XX
AC ACA18626;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #283.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
OS Escherichia coli.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen K, Zyskind JW;
PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX
WPI; 2003-023926/02.
DR P-PSDB; ABU14756.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 6496; 1766pp; English.
XX
PS The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of

```
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 417 BP; 87 A; 96 C; 116 G; 118 T; 0 U; 0 Other;

Query Match          34.8%; Score 417; DB 7; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.3e-127;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 GTGATTGCTTTGAAGTGTGAAGTGGCGAGCGCGCTTACTTGATTTGGCTGGGAATC 817
DB 1 GTGATTGCTTTGAAGTGTGAAGTGGCGAGCGCGCTTACTTGATTTGGCTGGGAATC 60

QY 818 CAGCAGTGGCGCGCGCTGGTGGCAATTGACCTTAATTCGCTGGCCTTACTCAATCGCGT 877
DB 61 CAGCAGTGGCGCGCGCTGGTGGCAATTGACCTTAATTCGCTGGCCTTACTCAATCGCGT 120

QY 878 CGACATTTGTTCCAGCGCGCAGTTTTGTGCAATCTCACCAATCCCAAAGATTTGTGTTT 937
DB 121 CGACATTTGTTCCAGCGCGCAGTTTTGTGCAATCTCACCAATCCCAAAGATTTGTGTTT 180

QY 938 CTGCGCGCGCTATTTCGCAATTCATATCGCCGACAGCGCACTGATGCAGTATATC 997
DB 181 CTGCGCGCGCTATTTCGCAATTCATATCGCCGACAGCGCACTGATGCAGTATATC 240

QY 998 GTGCTCGCGCTCACACACTATTGTGTCGATGATTTATTTGATGATCGGTACGCCACCTT 1057
DB 241 GTGCTCGCGCTCACACACTATTGTGTCGATGATTTATTTGATGATCGGTACGCCACCTT 300

QY 1058 GCTCAACGATTTGCTATGATTAAGGACCAAGCAGATGAAGCGCTGATTAAGATT 1117
DB 301 GCTCAACGATTTGCTATGATTAAGGACCAAGCAGATGAAGCGCTGATTAAGATT 360

QY 1118 TTCGGCTCGTGTGTTATGCTGCTGGAGCGCTGTAGCATCGCGAGGATCGGTGA 1174
DB 361 TTCGGCTCGTGTGTTATGCTGCTGGAGCGCTGTAGCATCGCGAGGATCGGTGA 417

RESULT 8
ACAS1728
ID ACAS1728 standard; DNA; 621 BP.
XX
AC ACAS1728;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #33385.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
OS Salmomella typhi.
XX
PN WO20027183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
```

```
XX
PA (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;
PI Wall D, Trawick JB, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR P-PSDB; ABU47858.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 39598; 1766pp; English.
```

```
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
```

Sequence 621 BP; 121 A; 170 C; 171 G; 159 T; 0 U; 0 Other;

Query Match 31.9%; Score 382.2; DB 7; Length 621;
Best Local Similarity 78.1%; Pred. No. 6.3e-116;
Matches 485; Conservative 0; Mismatches 133; Indels 3; Gaps 2;

```
QY 557 ATGACCTTAGAATGGTGGTTTGGCTACCTGCTGACATCGATCATTTTAACGCTGCGCA 616
DB 1 ATGACCTTTGAATGGTGGTTTGGCTATCTACTGACCTCAACTCTGTGAGTCTTCTCCG 60

QY 617 GGCTCTGGTCAATCAACACTATGACCACTCGCTCAACACCGGTTATCCGCGCG--GTG 674
DB 61 GGTTACGCGCCATCAATACCATGACGAGCTATCAACCATGATATCGTGGCGAGCG 120

QY 675 GCGTCTATTCTGGCTTTCAGACCGGACTGGCGATTATTTGCTGCTGGTGGCTGGGG 734
DB 121 GCTTCTATCGCGGGCTCCAGACCGGGCTGGGGATACATATCGTACTGGTGGCGCTCGGA 180

QY 735 TTGGG-ACGCTATTTTCCGCTCAGTGATTGGTTTGAAGTGTGGAAGTGGCGAGCGCG 793
DB 181 CTGGTACGCTCTTTTTCGCGCTCGCTCATCGCTTTTGAATTTCTGAATGGCTGGCGCG 240

QY 794 GCTTACTTGTATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGGCAATTAACCTTAA 853
DB 241 GCTTATCTTATCTGGTGGGTATCCAGCAATGGCGCGCGCGCTATCGATCTGCAT 300
```

QY 854 TCGCTGGCTCTACTCAATCGCTCGATTTGTTCCAGCGCGCAGTTTGTGAATCTC 913
CC
Db 301 ACTCTGCCCAAGCAATCGCGGGTGGCTGTTCATAACGGGCGATTTGTCAATCTA 360
QY 914 ACCAATCCCAAAAGATTTGTTTTCGGGGCGCTTATTTCCGCAATTCATCCCGCAA 973
CC
Db 361 ACCAATCCCAAAAGATTTGTTTTCGGGGCGCTTATTTCCGCAATTCATCCCGCAG 420
QY 974 CAGCCGCACTGATGAGTATATCGTCTCGCGTCACACTATTGTGTCGATATTATT 1033
CC
Db 421 CAACCCGCACTGGCGGAGTACTTATTCGGCGTCACAGATTTGTGATATGATT 480
QY 1034 GTGATGATCGTTTACCCACCTTCTCAACGGATTTGCTTATGATTAAGGACCAAG 1093
CC
Db 481 GTGATGATCGTTTACCCACCTTCTCAACGGATTTGCTTATGATTAAGGACCAAG 540
QY 1094 CAGATGAGCGCTGATTAAGATTTTCGCTCTGTTTATGCTGTTGAGCGCTGTTA 1153
CC
Db 541 CAGATGAGCGCTGATTAAGATTTTCGCTCTGTTTATGCTGTTGAGCGCTGTTA 600
QY 1154 GCATCGCGGAGGATCGTGA 1174
CC
Db 601 GCGTGGCAAGACCGGTGA 621

RESULT 9

ACA36103
ID ACA36103 standard; DNA; 618 BP.

ACA36103;

19-JUN-2003 (first entry)

Prokaryotic essential gene #17760.

Antisense; ds; prokaryotic essential gene; cell proliferation;
drug design; gene.

Klebsiella pneumoniae.

W0200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen K, Zyskind JW;

Wall D, Trawick JD, Carr G, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.

P-PSDB; ABU32233.

New antisense nucleic acids, useful for identifying proteins or screening

for homologous nucleic acids required for cellular proliferation to

isolate candidate molecules for rational drug discovery programs.

Claim 14; SEQ ID NO 23973; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of

the 6213 antisense sequences given in the specification where expression

of the nucleic acid inhibits proliferation of a cell. Also included are:

(1) a vector comprising a promoter operably linked to the nucleic acid

encoding a polypeptide whose expression is inhibited by the antisense

nucleic acid; (2) a host cell containing the vector; (3) an isolated

polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway,
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: the sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 618 BP; 108 A; 182 C; 179 G; 149 T; 0 U; 0 Other;

Query Match 31.2%; Score 375; DB 7; Length 618;

Best Local Similarity 77.6%; Pred. No. 1.6e-113;

Matches 479; Conservative 0; Mismatches 135; Indels 3; Gaps 2;

QY 557 ATGACCTTAGATGTTGTTGCTTACCTGCTGACATCGATCAATTTAAAGCTGTCCCA 616

Db 1 ATGACCATGATGTTGTTGCTTACCTGCTGACATCAATTTAAAGCTGTCCCG 60

QY 617 GGCCTGTGTGAATCAACACTATGACCACTCGCTGCTCAACCGTTA--TCCGCCCGGTG 674

Db 61 GGTTCGGAGCGATTAAATACCATGACCACTCCATTAAACCAAGGATACCGCGCGCG 120

QY 675 GGGCTATTGCTGGCTTCAGACCGGACTGGGATTCATATTGCTGTTGGCG-TGGG 733

Db 121 GCGTGAATGCGCGGTTCAGACCGGCTGGTTATTCAATCGTGTGTTGGGTGGT 180

QY 734 GTTGGGACGCTATTTTCCCGCTCAGTGAATGCTGTTGAAGTGTGAAGTGGCGAGGCGG 793

Db 181 CTCGCACTCTCTTCTCCCGCTCGGTGCTGGCTTTGAGGTGCTGAAATGGCGCGGCGC 240

QY 794 GCTTACTTGAATTTGGCTGGGAATCCAGCACTGGCGCGCGCTGTTGCAATGACCTTAA 853

Db 241 GCTACTGATCTGCTGGGTATCCAGCAATGGCGGGCGGAGGGGCGATCGACTGAAT 300

QY 854 TCGTGGCCTCTACTCAATCGGCTCGACATTTGTTCCAGCGCGCAGTTTGTGCAATCTC 913

Db 301 ACGTGGCGAAGCGCGAGCGCGCGGCAACTGTTCCAGCGCGGCTTCTGTAATCTC 360

QY 914 ACCAATCCCAAAAGATTTGTTTTCGGCGGCTATTTCCGCAATTCATCATCCGCAA 973

Db 361 ACCAATCCCAAAAGATTTGTTTTCGGCGGCTATTTTCGGCAGTTTATCTCCGCGAC 420

QY 974 CAGCCGCACTGATGAGTATATCGTCTCGCGCTCACCACTATTGTGTCGATATTATT 1033

Db 421 CAGCCGCGATGATGCAATATCTGTTCTGGGCTCACCACTCGTCTGATATCAT 480

QY 1034 GTGATGATCGTTTACGCCACCTTCTCAACGGAATGCTCTATGGAATTAAGGACCAAG 1093

Db 481 GTGATGATGCTATCGCACTCCCTCGCGAGGCGCATTTCCGATGGATCAAAAGGCCAAAG 540

QY 1094 CAGATGAGCGCTGAATTAAGATTTTCGCTCGTTTATGCTGTTGGGAGCGCTGTTA 1153

Db 541 CAGATGAGCGCTGAATTAAGATTTTCGCTCGTTTATGCTGTTGGGAGCGCTGCTC 600

QY 1154 GCATCGCGGAGGCGATGC 1170

Db 601 GCCTCCGCCGCCCATGTC 617

RESULT 10

ACA32157

ID ACA32157 standard; DNA; 618 BP.

AC ACA32157;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #13814.

DE Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Enterobacter cloacae.

OS WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI WPI; 2003-029326/02.

DR P-PSDB; ABU28287.

DR New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PT Claim 14; SEQ ID NO 20027; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

CC prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 618 BP; 106 A; 176 C; 183 G; 153 T; 0 U; 0 Other;

Query Match 29.8%; Score 357.4; DB 7; Length 618;

Best Local Similarity 75.9%; Pred. No. 1.1e-107;

Matches 468; Conservative 0; Mismatches 146; Indels 3; Gaps 2;

QY 557 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTACGGTGTGCGCA 616

Db 1 ATGACCTTCGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTACGGTGTGCGCG 60

QY 617 GGCTCTGCTCAATCAACACTATGACACCTGCTCAACACAGGTTA--TCCGGCGCGGTG 674

Db 61 GGCTCGGGGCTATTACACCATGACCATCTCCATCAATCACGGCTACCGGCTGGCGG 120

QY 675 GCGTCTATTCTGGGCTTCAGACCGGACTGGCGAATTCATATTGTGCTGTTGGCGT-GGG 733

Db 121 GCATCGATTGCGGCTTTCAGACCGGCTGGGTATTTCATATTGTGCTGCTGGGATCGGT 180

QY 734 GTTGGGACGCTATTTCCTCCGCTCAGTGAATTCGTTTGAAGTGTGAAGTGGCGAGCGGG 793

Db 181 CTGGGACCTCTGTTCTCCCGCTCCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240

QY 794 GCTTACTTGAATTTGGCTGGGAATCCAGACGTGGCGCGCGCTGGTGCATTAATGACCTTAA 853

Db 241 GCGTATCTGATTTGGCTCGGCATCCAGACGTGGCGCGCGCTCCATCAACCTGAAT 300

QY 854 TCGTGGCCTCTACTCAATCGCTGCGACATTTGTTCCAGCGGCGAGTTCCTGGAATCTC 913

Db 301 ACGTGTGCTCTGACGCAAAACCGCGGCTCTCTGTTAAGCGTGGGTATTCGTCACCTG 360

QY 914 ACCAATCCCAAAAGTATTTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATCGCGCAA 973

Db 361 ACCAACCGAAGACATCGTTCCTCGCGGCGCTGTTCCCGCAGTTTATCGTTCCGCTAT 420

QY 974 CAGCGCAATGATGACGATATATCGTCTGGCGTCAACCACTATTCGTCGATATTT 1033

Db 421 CAGCTCAGGTGATGACGATGACGTTGGTGGCGGCGCACTATTCGTCGATATTCATC 480

QY 1034 GTGATGATCGGTTACGCCACCTTGTCTCAACGATTCGCTCTATGATTTAAAGGACCAAG 1093

Db 481 GTGATGATGTTACGGACGCTGGCGGCGGATTTAAAGGCGCTAAG 540

QY 1094 CAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGCTGGAGCGCTGTTA 1153

Db 541 CAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGCTGGAGCGCTGCTT 600

QY 1154 GCATCGGCGAGGATGC 1170

Db 601 CGTTCAGCGGCTACGC 617

RESULT 11

ACA48809

ID ACA48809 standard; DNA; 623 BP.

XX ACA48809;

XX 19-JUN-2003 (first entry)

DT Prokaryotic essential gene #30465.

DE Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Salmonella paratyphi.

OS WO200277183-A2.

XX 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI: 2003-029926/02.
DR P-PSDB; ABU44939.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 36679; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 623 BP; 124 A; 171 C; 168 G; 160 T; 0 U; 0 Other;
Query Match 29.0%; Score 348.2; DB 7; Length 623;
Best Local Similarity 76.6%; Pred. No. 1.3e-104;
Matches 479; Conservative 0; Mismatches 138; Indels 8; Gaps 4;
QY 557 ATGACCTTAGAATGGTGGTTGGTTCCTACCTCGTGCATCATGATCATTTTAACTGTCGCCA 616
DB 1 ATGACCTTTGAATGGTGGTTGGTTCCTATCTACTGCACTCAACTCTGCTGAGTCTTTCTCCG 60
QY 617 GGCTCTGGTGCATACACTATGACCACTCGCTCAACCCAGGTTATCCGGCGG--GTG 674
DB 61 GGTTCAGGCCCAATCAATACCATGACGAGCTTATCAACCATGGATATCGTGGCGACGC 120
QY 675 GCCTCTATTGCTGGGTTTCAGACCGGACTGGCGATTTCATTTGTGCTGGTTGGCGTGGGG 734
DB 121 GCTTCTATCGCGGACTCCAGACCGGACTGGGATACATATGCTACTGCTGGCGGCTCGGA 180
QY 735 TTGGG-ACGCTATTTCCCGCTCAGTGATGGCTTTGAAGTCTTGAAGTGGCGACGGCGG 793

DB 181 CTGGGTACGCTCTTTTCGGCGCTCGCTCTCTCTTTTGAATTTCTGAATGGGCTGGCGG 240
QY 794 GCTTACTTGAATTTGGCTGGGAATCCAGCACTGGCGGCCCTGGTGCATTAACCTTAA 853
DB 241 GCTTATCTTATCTGGCTGGGTATCCAGCAATGGCG--CGCAGGCGCTATCGATCTGCAT 297
QY 854 TCGTGGGCTCTACTCAATCGGTCGACATTTGTCAGCGCGCAGTTTTTGTGAATCTC 913
DB 298 ACTTCGCGCCAGACGCAATCGGGGTCGGCTGTTCAACGGGCGATTTTGTCAATCTA 357
QY 914 ACCAATCCAAAAGTATTTGTTTTCGGCGGCGTATTTCCGCAATTCATCATGCGGCAA 973
DB 358 ACCAATCCAAAAGTATTTGTTTTCGGCGGCGCTGTTTCCGCAATTCATCATGCGGCG 417
QY 974 CAGCGGCACTGATGCAATATATCGTCTGGGCTGACCACTATTGTTGGTGCATATTATT 1033
DB 418 CAACCGCAACTGGCGGAGTACCTCATTTCTGGGCTACACAGATTGTTGGTGAATGATT 477
QY 1034 GTGATGATCGGTTACGCCACCCCTTGCTCAACGG--ATTGCTCTATGATTAAGGACCAA 1091
DB 478 GTGATGACCGGTTACGCCACACTGCGCTAGCGCACTTGCGGCTGATTTAAAGGACCAA 537
QY 1092 AGCAGATCAAGCGCTGATTAAGATTTTCGCTCGTGTGTTTATGCTGGTGGAGCGCTGT 1151
DB 538 AGCAGATCAAGCGCTGATTAAGCGTTTGGTTCGTTGTTTATGCTGCTAGGCGCGCTCC 597
QY 1152 TAGCATCGCGGAGGCAATCGCTGAA 1176
DB 598 TGGCGTCGCGAAGACACGCGTGATA 622
RESULT 12
ACA19693/c
ID ACA19693 standard; DNA; 1017 BP.
XX
AC ACA19693;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #1350.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Salmonella typhimurium.
XX
PN WC200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI: 2003-029926/02.
DR P-PSDB; ABU15823.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 7563; 1766pp; English.
XX

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation or to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1017 BP; 208 A; 251 C; 305 G; 253 T; 0 U; 0 Other;

Query Match 24.9%; Score 298.8; DB 7; Length 1017;

Best Local Similarity 79.4%; Pred. No. 4.6e-88;

Matches 354; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1 AGAATAATGTGGAGATCGCAGCGCCATCGAATGTGCGATATATAGCGTTACGCCAC 60
Db 446 AGGAACAGGGTGGCTATCGCGGCCCCATTTAGTGTGCGCAGGATATACGTTACGCCAC 387
QY 61 GGACCGGGTGAACCTTCCTGCTGCGCAGATCGCCAGATCATCAACATATCATTAAG 120
Db 386 GGCCCGGGTCAATTTCTCTGCTGCGCAGACGCGCTTAAGTCTCAACATATCGTTAAA 327
QY 121 CGATTAAATGCCCGAGATCGCGATCGGCTAACAGCGCAGCGAACGTCCTGCCCGCGA 180
Db 326 TGATCGACATGACCCCGATGCGGTCCGATACATCCGCCGAGCGCCCTTGACACGG 267
QY 181 TGCTCGATGATTAAGCATCAACCCCAATGGAACAGGTATAGGCCAGTTCGCGATAT 240
Db 266 TGCTCGATGATGAAGATATCAAAACCGAGATGAAAAGGTATATAGCGCAGTTCAGCGTAT 207
QY 241 TTTCAGTGTCTCTCAATAGCCCGCGGAGATGATACACCCGCTCATGTGTGTGTGCG 300
Db 206 TTTCATAACTCTCAATGCGCGCAGGACAAATGACATGCTCGATCGTTGGAATCATTT 147
QY 301 CGAAACGGAAGAGCGACCGGATGTCTATCCACACAGTAATCTGTCTTATACCGC 360
Db 146 CGAAACGGAAGAGCGACCGGATGTGTACCCAGCGCGATAAATCTGCGCTCTTCCCGC 87
QY 361 TGACCGCAGAAATCAGTCAAGCGGTCCCATGTGTAAAGCAGCAACGCGTTTCTTGTGT 420
Db 86 TGACCGCAGAAATCGGTCAAGCGGTCCCATGTGTAAAGCAGCAACGCGTTTCTTGTGT 27
QY 421 TCCAGTCTTTTGTGCTGTAACAT 446
Db 26 TCCAGTCTTTTGTGCTGTAACAT 1

RESULT 13

ACA51727/c

ID ACA51727 standard; DNA; 1017 BP.
XX ACA51727;
AC 19-JUN-2003 (first entry)
XX 19-JUN-2003 (first entry)
DT Prokaryotic essential gene #33384.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Salmomella typhi.
OS WO200277183-A2.
PN 03-OCT-2002.
PD 21-MAR-2002; 2002WO-US009107.
PF 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR P-PSDB; ABU47857.
DR New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 14; SEQ ID NO 39597; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation or to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1017 BP; 208 A; 249 C; 305 G; 255 T; 0 U; 0 Other;

Query Match 24.8%; Score 297.2; DB 7; Length 1017;
Best Local Similarity 79.1%; Pred. No. 1.6e-87;
Matches 353; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
Qy 1 AGAAATATGTGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 446 AGAACAGGGTGGCTATCGCGGCCCCCATTTAGTGGCCAGGATATACGTTACGCCAC 387
Qy 61 GGACCGGGTGAACCTCTGCTGCCAGATGCGCGCAGATCATCAACATATCAATTAAG 120
Db 386 GACCCCGGGTCAATTTCCCTGCTGCCAGAACGCGGCTAAGTCTCAACATAATCGTTAAA 327
Qy 121 CGATTAACTGCCGAGATCGCGATCGGCTAACAGCGACCGGAACGTCCTGCGCCGGA 180
Db 326 TGATCGACATGACCCCGATGCGGGTCCGATACATCCGCCGAGCGCCTTGACCGG 267
Qy 181 TGTGATGATTAAGACATCAACCCCAATGGAACAGGTCTATAGGCCAGTTCGCGATAT 240
Db 266 TGTGATGATGAAGATATCAAAACCGAGATGAAAGGTCTATAGCCAGTTCAGCGTAT 207
Qy 241 TTTACGTAGTCTCAATACGCGCGCGCATGACTACACCCGGTCAATGCTGTGTGG 300
Db 206 TTTACATACTCTCAATGCGGCGAGCAAAATGACGATCGTCCGATCGTTGGAGTCATTC 147
Qy 301 CGAAACGGACAAAGCGCACCGGAATGTCTATCCACACAGTAAATCTGTCTCATACGC 360
Db 146 CGAAACGGACAAAGCGCACCGGAATGTCTATCCACCGCGATAAATCTCGGCTCTTCGCGC 87
Qy 361 TGACGCCAGAAATCAGTCAAGCGTCCCATGTGTAAGACAGAGAAACGCGTTTCTCTTGT 420
Db 86 TGACGCCAAATTCGGTCAAGCGGCCCATAGCAAAAGCGGCAACGCAATTTCTCTTGT 27
Qy 421 TCCAGTCTTTTGTGCTGCTGAACAT 446
Db 26 TCCAGTCAATTTGCTGCTGMAACAT 1

RESULT 14
AAS88021/c
ID AAS88021 standard; cDNA; 2054 BP.
XX AC AAS88021;
XX DT 13-FEB-2002 (first entry)
XX DE DNA encoding novel human diagnostic protein #23825.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX DE food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX FN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX FA (HYSE-) HYSEQ INC.
XX FI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-FSDB; AEG23834.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.

PS Claim 1; SEQ ID NO 23825; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2054 BP; 477 A; 527 C; 563 G; 487 T; 0 U; 0 Other;
Query Match 24.3%; Score 291.4; DB 5; Length 2054;
Best Local Similarity 99.7%; Pred. No. 2e-85;
Matches 292; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AGAAATATGTGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 460 AGAAATATGTGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 401
Qy 61 GGACCGGGTGAACCTCTGCTGCCAGATGCGCGCAGATCATCAACATATCAATTAAG 120
Db 400 GGACCGGGTGAACCTCTGCTGCCAGATGCGCGCAGATCATCAACATATCAATTAAG 341
Qy 121 CGATTAACTGCCGAGATGCGGATCGGCTAACAGCGACCGGACGTCCTGCGCCGGA 180
Db 340 CGATTAACTGCCGAGATGCGGATCGGCTAACAGCGACCGGACGTCCTGCGCCGGA 281
Qy 181 TGTGCTGATGATTAGACATCAAAACCCCAATGGAACAGGTCATAGGCCAGTTCGCGCATAT 240
Db 280 TGTGCTGATGATTAGACATCAAAACCCCAATGGAACAGGTCATAGGCCAGTTCGCGCATAT 221
Qy 241 TTTACGTAGTCTCAATAGCGCGCGGCGAGATGACTACCAACCGGTCATGGTG 293
Db 220 TTTACGTAGTCTCAATAGCGCGCGGCGAGATGACTACCAACCGGTCATGGTG 168
RESULT 15
ACA32154/c
ID ACA32154 standard; DNA; 990 BP.
XX AC ACA32154;
XX DT 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #13811.
XX DE Antisense; ds; prokaryotic essential gene; cell proliferation;
XX DE drug design; gene.
XX OS Enterobacter cloacae.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX XX 21-MAR-2002; 2002WO-US009107.
XX XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX PA
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu RH;
XX
XX WPI; 2003-029926/02.
DR F-PSDB; ABU28284.
XX

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 20024; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: the sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 990 BP; 187 A; 266 C; 309 G; 228 T; 0 U; 0 Other;

Query Match 24.2%; Score 290.8; DB 7; Length 990;

Best Local Similarity 78.3%; Pred. No. 2.1e-85;

Matches 349; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1 AGAATATATGTGGAGATCGACCGCCCATCGAATGTCCAGTATATAGCGTTTACGCCAC 60
DB 446 AAAAAACAGCGTGAATGGCGCGCCCATGGAGTGGCGCAAGATATACCGCTTGCGCCAG 387
QY 61 GGACCGGGCTGAACCTCTGTGTCAGAAATGCCGCCAGATCATCAACATATATTAAAG 120
DB 386 GGGCGGGCTGAACCTCTGTGTCAGAAATGCCGCCAGATCATCGTGAATCGCTGAAG 327
QY 121 CGATTAAATGCCCGAGATCGGATCGGCTAACAGGGACCGGAAACGTCCTCCCGCGGA 180
DB 326 CGATCGAGTGGCCCGACGGTGTGTCCGGCAGCATACGTCGGGACAGCCCTGTCCACGA 267
QY 181 TGGTCGATGATTAGACATCAAAACCCCAATGGAACAGGTTCATAGGCGAGTTCGGCATAT 240
DB 266 TGGTCGATGATCAGACATCGNACCCCATATGGACAGGTCATAGGCGAGCTCGGGTAT 207

QY 241 TTTAGTAGTCTCTCAATACGCGCCCGGCAGATGACTACACCGGTCTATGCTGTGGC 300
DB 206 TTAACGTAGCTTTTCAATACGCGCGGGGCAGACCAATACCCGGTCAATTTTTCATTT 147
QY 301 CGAAAAACGACAAAGCGCACCGGAATGTCTATCCACACACAGTAAACTCTGCTTCATCACGC 360
DB 146 CGGAACCGAACGAAACGACCGGGATGCCCCCGACGCGCTAAACTCATCTTCTCACGC 87
QY 361 TGACGCCAGAAATCAGTACAGCGGTCCCATGTGTAAGCAGCAAAACGGTTCCTTGT 420
DB 86 TGCCGCCAGAAATCGGTACGCGGCCCATAGAGAAAGCAGCAAAAGCGTTTCTCTGT 27
QY 421 TCCAGTCTTTTCTGCTGTAACAT 446
DB 26 TCCAGTCTTTTCTGCTGTAACAT 1

Search completed: March 28, 2004, 14:37:59
Job time : 603.686 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:10:49 ; Search time 120.331 Seconds
(without alignments)
5534.263 Million cell updates/sec

Title: US-09-847-392-1

Perfect score: 1200
Sequence: 1 agaaataatgtgagatcgc.....atctcgatgcgcgtaaac 1200

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1200	100.0	1200	4	US-09-396-357-1
2	417	34.8	417	4	US-09-711-164-274
3	391.8	32.6	645	4	US-09-489-039A-895
C 4	297.2	24.8	1041	4	US-09-489-039A-851
C 5	207.4	17.3	263	4	US-09-711-164-107
C 6	144.8	12.1	1038	4	US-09-543-681A-238
C 7	112.4	9.4	834	4	US-09-252-991A-6447
C 8	102.4	8.5	930	4	US-09-252-991A-6103
9	75.6	5.4	663	4	US-09-328-352-777
10	59	4.9	645	4	US-09-489-039A-2622
11	46	3.8	1830121	4	US-09-557-884-1
12	46	3.8	1830121	4	US-09-643-990A-1
13	37.2	3.1	633	4	US-09-328-352-2682
14	36	3.0	312	4	US-09-543-681A-3779
15	35.4	2.9	1028	4	US-09-655-908-17
C 16	34.4	2.9	612	4	US-09-489-039A-1357
C 17	34.4	2.9	1431	4	US-09-489-039A-1316
C 18	33.8	2.8	3534	4	US-09-134-001C-2269
C 19	33.2	2.8	789	4	US-09-252-991A-2459
20	33.2	2.8	1296	4	US-09-252-991A-2216
21	33.2	2.8	1941	4	US-09-252-991A-2378
C 22	33.2	2.8	2820	4	US-09-252-991A-2538
C 23	33.2	2.8	18993	4	US-09-961-827-15
C 24	33.2	2.8	4403765	3	US-09-103-840A-2
C 25	33.2	2.8	4411529	3	US-09-103-840A-1
26	32.8	2.7	1442	4	US-09-634-238-82
C 27	32.8	2.7	16075	3	US-09-096-867-1

C	28	32.8	2.7	16075	3	US-09-096-867-1	Sequence 1, Appli
	29	32.6	2.7	968	4	US-09-023-655-1112	Sequence 1112, Ap
	30	32.4	2.7	474	4	US-09-621-976-18033	Sequence 18033, A
C	31	32.4	2.7	161652	4	US-09-497-855A-40	Sequence 40, Appl
	32	32.4	2.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	33	32.4	2.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	34	32.2	2.7	681	4	US-09-252-991A-15039	Sequence 15039, A
C	35	32	2.7	1001	4	US-09-641-638-278	Sequence 278, App
C	36	32	2.7	1597	4	US-09-634-238-195	Sequence 195, App
C	37	31.8	2.6	738	4	US-09-252-991A-6015	Sequence 6015, Ap
	38	31.8	2.6	1128	4	US-09-252-991A-6376	Sequence 6376, Ap
	39	31.8	2.6	2136	4	US-09-252-991A-6294	Sequence 6294, Ap
C	40	31.6	2.6	3948	4	US-09-252-991A-6175	Sequence 6175, Ap
	41	31.6	2.6	25356	4	US-09-976-594-750	Sequence 750, App
	42	31.4	2.6	455	4	US-09-621-976-9043	Sequence 9043, Ap
C	43	31.4	2.6	13987	2	US-08-804-227C-13	Sequence 13, Appl
C	44	31.4	2.6	44377	2	US-08-804-227C-7	Sequence 7, Appli
C	45	31.4	2.6	44377	2	US-08-804-198-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-396-357-1
; Sequence 1, Application US/09396357
; Patent No. 6303348
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKHAROVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENAMIOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHMAKOVA, IRINA L'OVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFRS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/396,357
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: RU98118425
; EARLIER FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (557)..(1171)
US-09-396-357-1

Query Match	100.0%	Score 1200;	DB 4;	Length 1200;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1200;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
QY	1	AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC	60	
DB	1	AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC	60	
QY	61	GGACCGGGCTGACCTCTGCTGCCAGATGCCCGCAGATCATCAACATCAATTAAG	120	
DB	61	GGACCGGGCTGACCTCTGCTGCCAGATGCCCGCAGATCATCAACATCAATTAAG	120	
QY	121	CGATTAACATGCCCCGAGATCGCGTAAACAGCGACCGGACGTCCTCGCCCGCGA	180	
DB	121	CGATTAACATGCCCCGAGATCGCGTAAACAGCGACCGGACGTCCTCGCCCGCGA	180	
QY	181	TGTCGATGATTAAGACATCAACCCCAATGGAAGAGTCTAGCCGAGTTCGGCATAT	240	
DB	181	TGTCGATGATTAAGACATCAACCCCAATGGAAGAGTCTAGCCGAGTTCGGCATAT	240	
QY	241	TTTACGTAGTCTCAATACCGCCCGGCGAGATGACTACCACCGGTCGTGCTGTGG	300	

533 AAATAGACACCGGGATTTCATCATGACCTTAGAATGGTGGTTGGCTACCTGCTGACA 592
1 AAACAGACAATCAGGGAGTTTCTCATGACCAATTCAGTGGTGGTTGGCTACCTGCTGACA 60
593 TCGATCATTTTAAACCTCTGCCAGGCTCTGGTCAATCAACATCACTAGACCACTCGCTC 652
61 TCAATATCTTCAGCTCTGCCCGGTTCCGAGCGAATTAATCAATGACCACTCATTT 120
653 AACCAAGGTTA - TCCGCGCGGTGGCGTCTATCTGGGTTCAGACCGGACATGGCGATT 710
121 AACCAAGGATACCGCGCGCGCGCTGATTCGCGGTTTCAGAGTGGGCTGGCGATT 180
711 CATATGTGCTGGTGGCG - TGGGGTTGGGACGCTATTTCCGCTCAGATGATGGCTTT 769
181 CATATGTGCTGGTGGCGTCTGGTCTGGGACTCTCTCTCCGCTCGGTGGCGCTTT 240
770 GAAGTGTGAAGTGGGCGCGCGCTTACTTGTATTTGGCTGGGAATCCAGCAGTGGCGC 829
241 GAGGTGCTGAATGGCGCGCGCGCTTACTTGTATTTGGCTGGGAATCCAGCAGTGGCGC 300
830 GCCGTGGTGAATGACCTTAATTCGCTGGCTCTACTCAATCGGTGACATTTGTTTC 889
301 GCGGCGAGGCGGATCGCTGAATACGCTGGCGAAGCGGAGCGCGCGCAACTGTTT 360
890 CAGCGCGAGTGTGTGAATCTCACCATCCCAAAAGTATTTGTTCTGGCGGCGCTA 949
361 CAGCGCGGTTGCTGTAATCTCACCATCCCAAAAGTATTTGTTCTGGCGGCGCTG 420
950 TTTCGCGAATTCATATCGCGCAACAGCGGCAATGATGAGTATATCGTCTGGCGCTC 1009
421 TTTCGCGAGTTTATCTCCCGCAGCCAGCGGAGGTGATGATATATATCGTCTGGCGCTC 480
1010 ACCATTTGCTGCTGATATTTATTTGATGATGCTTACGCCACCTTGTCAACGATT 1069
481 ACCATCTGCTGCTGATATCAATGATGATGATGATGATGATGATGATGATGATGAT 540
1070 GCTATGATTAAGGACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1129
541 TCCGATGATCAAAAGGCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 600
1130 TTATGCTGGTGGAGCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 1174
601 TTATGCTGGTGGAGCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 645

RESULT 4
US-09-489-039A-851/c
; Sequence 851, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.200401
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 851
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-851

Query Match 24.8%; Score 297.2; DB 4; Length 1041;
Best Local Similarity 75.7%; Pred. No. 3.1e-88;
Matches 368; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

1 AGAATAATGTGAGATCGCACCGCGCTGATGATGATGATGATGATGATGATGATGATGAT 60
494 AGAAAAGCGTGGCAATGCTCCGCCCATAGATGAGCAAGATAACCGCTTTTCGCCAG 435

61 GGACCGGGCTGAACCTCTCTGCTGCGAGAAATGCGCGCAGATCATCAATATCAITTAAG 120
434 TGTCCAGGAACCACTCTCTGCTGCGAGAGCGCGCGCAGGTTCATCAATAGTCGTGAAG 375
121 CGATTAACATGCGCGAGATCGGATCGGCTTAAAGCGGACCGGAAAGTCCCTGCCCGCGA 180
374 TTAACACGCTGTCGCGGTGGTATCTGAGAGCAGGCGCGGAGGCTCCCTGCCCGCGA 315
181 TGGTCGATGATTAAGACATCAAAACCCCAATGAAACAGGTTCATAGCCGAGTTCGCGATAT 240
314 TGGTCAATGATCATCATGCTCAAAAGCGCGAGTGAAGAGATCGTAGGCCACTTCAGCATAC 255
241 TTTACGTAGTCTCAATACCGCGCGCGCAGATCACTACACCGGCTCATGCTGCTGCG 300
254 TTACGTAGTCTCAATGCGCGCGCGGCAATCAACACGAGATCGTATGTTGCGCG 195
301 CGAAAACGGAACAAAGCGCACCGGAATGTCAATCCACACAGTAAATCTCTGCTTCATCAGC 360
194 CAAAAGCGGACAAAGCGTACCGGTACATCGTCGACGCTTTAAATTCAGCTTCTCGCGC 135
361 TGACCGCAGAAATCAGTCAGGCTCCCATGTTAAAGCAGCAACCGCTTTCTCTGTT 420
134 TGTCCGAGAAATCCGTCAGCGCGCGCATCTAAAGCGCGCAACCGCTTTCTCTGTT 75
421 TCCAGTCTTTTCTGCTGAAACATCGGTAATCTGCTCTTTAAACACGATAAATCGT 480
74 TCCAGTCTTTTCTGCTGAAACATCGGTAATCTGCTCTTTAAACCGTGAATCGT 15
481 TTTTCT 486
14 TTTTCT 9

RESULT 5
US-09-711-164-107/c
; Sequence 107, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERE
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-711-164-107

Query Match 17.3%; Score 207.4; DB 4; Length 263;
Best Local Similarity 94.3%; Pred. No. 8.8e-59;
Matches 248; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

861 CCTCTACTCAATCGCTGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACAATC 920
263 CCCCTACTCAATCGCTGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACAATC 204
921 CCAAAGTATTTGTTTCTGCGCGCGCTATTTCCGCAATTCATCATGCGCGCAACGCGC 980
203 CCAAAGTATTTGTTTCTGCGCGCGCTATTTCCGCAATTCATCATGCGCGCAACGCGC 144
981 AACTGATGCGATATATCGTCTC - GGCGTCACCACTATTGTTGTCGATATTTATGATG 1039
143 AACTGATGCGATATATCGTCTC - GGCGTCACCACTATTGTTGTCGATATTTATGATG 84
1040 ATC - GGTTACGCCACCTTCTCTACGCGATTG - CTCATGATTTAAGCAACCAAGCAG 1096

Db 83 ATCGGTTAGCCACCCCTTCTCAACGGATGGCTCTAAGGGATTAAAGGCCCAACAA 24
QY 1097 ATGAAGCGCTGTAATAGATTTT 1119
Db 23 ATGAAGCGCTGTAATAGATTTT 1

RESULT 6

US-09-543-681A-238/c
; Sequence 238, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 238
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-238

Query Match 12.1%; Score 144.8; DB 4; Length 1038;
Best Local Similarity 59.5%; Pred. No. 1.2e-37;
Matches 245; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
QY 14 AGATCGCACCGCCCATCGAATGCGCAGTATATAGGTTTACGCCACGGACCGGGCTGAA 73
Db 472 AATAGCGCGCCCATCGAATGCGCTAACGCAAAATAGTGTGATAGTAAAGCAAAA 413
QY 74 CTTCTCTGCCAGATGCCCGCAGATCATCAACATATATCAATTAAGCGATTAACATGCC 133
Db 412 CGATAGTATCAATAAATGTCGATAATCATACATCAATCAATCAATCAATCAATCAAT 353
QY 134 CGAGATGCGGATCGGCTAACAGCGCACCGGACGTCCTCCCGCGGATGTCGATGATTA 193
Db 352 CTTTTTGGGATCATTAATAGCGCCCTGTAATCTTCCCGCGATGATCGAGAGAA 293
QY 194 AGACATCAAAACCCCAATGAACAGATCATAGCGAGTTTCGCAATTTTACGTAGTCT 253
Db 292 AATATATCAATCAACCAAGTGATATAAATCGAAGGCCACATTCAGGGTATTTTACATAACTTT 233
QY 254 CAATACGCCCGCGGACATCACTACACCGGTCATGCTGTGCGGGAACGGCAA 313
Db 232 CACTACGGCCAGACGATCACCAACGCTTTTGTGTGATGAAGAGCGCCATGAACAT 173
QY 314 AGCGCACCGGAATGTATCCACACCAAGTAACTCTGCTTCATCAGCTGACGCCAGAAAT 373
Db 172 AGCGAATGCTACGCCATCGACTCCGGTAAATTCGCTTCTGCTGTTGGTTCCAAAAGT 113
QY 374 CAGTCACGGTCCCATGTTAAAGAGCAACACGGTTTCTCTTTTCCCA 425
Db 112 CCAATAGATACCATTCGTTAAAGCAGAGATTTGCTTTTTCACGAGATAACCA 61

RESULT 7

US-09-252-991A-6447
; Sequence 6447, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6447
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6447

Query Match 9.4%; Score 112.4; DB 4; Length 834;
Best Local Similarity 53.5%; Pred. No. 6.3e-27;
Matches 324; Conservative 0; Mismatches 276; Indels 6; Gaps 4;
QY 572 TGGTTTGCTACCTGCTACATCATTTACGCTGCGCAGGCTCTGTGCAATC 631
Db 220 TGGTTGCGCTTCTTCTGCTGCTGCGCCATCAGCTGCGCGCGCGCGGAGCCATC 279
QY 632 AACACTATGACACCTCGCTCAACACAGGTTATCCGCGCGGTGC--GTCTATTGTGG 699
Db 280 GCCTCGATGCTCTCGGGTTTCAGTACGGCTTTGCGGAGGCTACTGGAACGGCTGGC 339
QY 690 CTTGACACCGACTGGCGATTATTTGCTGTGTGGCG--TGGGTTGGAGCTATTT 748
Db 340 CTSAGATCGGCTGCGCTCGCATTCGCCATTTCGCCCGCGCGCTCGGTGCGTCTG 399
QY 749 TCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGCGAGCGCGCTTACTTGAATTGG 808
Db 400 GCGACTCGGCACCTGGCTTCAGCTGATCAAGTGGTTTCGGCGTGGCTACCTGGTGTAC 459
QY 809 CTGGAAATCCAGCAGTGGC--GCGCGCTGGTGCAATGACCTTAATTCGCTGGCCTCTA 866
Db 460 CTGGCGGTGCGCGCAGTGGCGCGCGCCACAGCGCTTGAGCACCGATGGCGAACGGCCT 519
QY 867 CCAATCGCTCGACATT--TGTTCAGCGCGCAGTTTGTGAATCTCACCAATCCCAA 925
Db 520 CTGGGCGAACGTTGACCTGCTGTGCTGGTGTCTGTCAACGCGCAGCATCCCAAG 579
QY 926 AGTATGTGTTTTCGGCGCGCTATTTCCGCAATTCATTCGCGCAACACCGCAACTG 985
Db 580 GCGGTGATCTTCATGCTCGCGTGTGCTCGCAGTTTCATCGACCGCGCACCGCGCTCTG 639
QY 986 ATGCAATATTCGCTCGCGGTCCACCATTTCTGTGCTGATATTTTGTGATGATCGGT 1045
Db 640 GCGCAATACCTGATCATGCGCGCGCACCATGATCTGCTGACCTGATCTGATGCGCGC 699
QY 1046 TACGCCACCTTGTCTCAACGATTCCTTATGATTAAGAACCAAGACGATGAAGCG 1105
Db 700 TACACCGGCTGGCTGCGCGTGTCTACGCTACTGCTGCTGCGCGCGCAGCAGAGCTG 759
QY 1106 CTGAATAGATTTTCGGCTGCTGTTTATGCTGTGGAGCGCTGTAGCATCGCGAGG 1165
Db 760 GTGAACCGTACCTTCGCGAGCTGTTCGTGCTGCCCGGGGCTGCTGGCAGCGTACGC 819
QY 1166 CATGCG 1171
Db 820 CGAGCG 825

RESULT 8

US-09-252-991A-6103/c
; Sequence 6103, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

QY 910 TCTACCAATCCAAAGTATTGTGTTCTGCGCGCTATTTCCGCAATTCATCATGCC 969
Db 384 TGTGCTTACCCCAAAATGGCGTCTTCTACGTTTCGTTTATACCTCAGTTTATCCCGC 443
QY 970 GCAACAGCGGAACATGATGAGTATATCGTGTGCGCTACCACTATTGTGTCGATAT 1029
Db 444 CGGGCAATCCACCGTGAGTGTGACTTTTCTCTGCTGCTATCATGATTTGATTTGATC 503
QY 1030 TATTGTGATGATCGGTTAGCCACCTTCTGCTCAACGGATTCTCTATGATTTAAAGGACC 1089
Db 504 GCTGTGCTGTTAAGCTGATTACCGCCACTCGCTATGCCGCGAGCATCTGAAACACC 563
QY 1090 AAGCAGATGAAGCGCTGAATAAGATTTCGCTCGTTTATGCTGTGGAGCGCT 1149
Db 564 AGCCGTAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 623
QY 1150 GTTAGCA 1156
Db 624 GCTGGCA 630

RESULT 11
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 3.8%; Score 46; DB 4; Length 1830121;
Best Local Similarity 52.5%; Pred. No. 0.0077;
Matches 128; Conservative 0; Mismatches 110; Indels 6; Gaps 1;
QY 1 AGAATAATGTGAGATCGCACCGCCCATCGCAATGTGCCAGTATATACCGTTAGCCAC 60

Db 689153 AATAATAGTGTCAATCAATGCGCCCATTAATGAATGACGAAGCAAAATGCTGCTGTAATAG 689212
QY 61 GGACGGGCTGAACCTCCTGCTGCCAGATCGCCGAGATCATCAACATAATCATTTAAAG 120
Db 689213 CTGAAGAGTGGGTACTTTTTCGATGATTTTTCGCATATCAATCGACATAGAAACGAAT 689272
QY 121 CGATTACATGCCCGAGATCGGATCGGCTAACAGGCGACCGGAAAGTCCCTGCCCGCA 180
Db 689273 TCATCTAAATGCCCTTTTGGAGG-----ATAATACGCTGTGAATAGCTTTGACCGCGA 689326
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCATAGGCCCATTTCCGCATAT 240
Db 689327 TGATCAATAGCATACATCGTAACCTTTGATGATAAAATCATAAGCGAGCTCTGACCAT 689386
QY 241 TTTA 244
Db 689387 TTCA 689390

RESULT 12
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 3.8%; Score 46; DB 4; Length 1830121;
Best Local Similarity 52.5%; Pred. No. 0.0077;
Matches 128; Conservative 0; Mismatches 110; Indels 6; Gaps 1;

```
QY 1 AGAATAATGTGAGATCGCAACGCCATCGAATGTGCCAGTATATAGCGTTIACGCCAC 60
Db 689153 AAATAATAGTTGCAATCAATCGCCCATTAATGAGCAACAATGCTGGTGGTGAATAG 689212
QY 61 GGACCGGCTGAACCTCTGCTGCCAGATGCGCGCAGATCATCAACATATCATTAAG 120
Db 689213 CTGAAGAAGTGGGTACTTTTTCGATGATTTTCGCCATATCATCGACATAGAAACGAAAT 689272
QY 121 CGATTAACATGCGCGAGATGCGGATCGGCTAACAGCGCACCGGACGTCCTGCCCGCA 180
Db 689273 TCATCTAAATGCCCTTTTGGAG-----ATAATACGCTGTGATAGCCTTGACCGCA 689326
QY 181 TGGTCATGATTAAGACATCAACCCCAATGGAACAGGTGATAGGCGAGTTCGCGATAT 240
Db 689327 TGAATCAATAGCAATACATCGTAACCTTGATGATAAAATCATTAAGCGAGCTGTGACCAT 689386
QY 241 TTATA 244
Db 689387 TTCA 689390

RESULT 13
US-09-328-352-2682
; Sequence 2682, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 3252
; SEQ ID NO 2682
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2682

Query Match 3.1%; Score 37.2; DB 4; Length 633;
Best Local Similarity 46.2%; Pred. No. 0.05;
Matches 123; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 880 ACATTGTTCCAGCGCGAGTTTGTGAATCTCACCATCCCAAAAGTATTGTTTCT 939
Db 348 AAAGCTTTTCTATCAAGGTTGTTTGCTCTACTTAATCCAAAACAATTGCTTTT 407
QY 940 GCGGCGCTATTTCGCAATTCATCATCGCGCAACAGCGGCAACTGATGATATCGT 999
Db 408 TATTGCCCTTTTCCACAGTTTATTGATATTAATAAAGAAATATTAATCAGTTTGT 467
QY 1000 GCTCGGCGTCAACACTATTGTGCTGATATTATTGTGATCGTTACGCCACCTTGC 1059
Db 468 TTATCTTTTAACTTCTGCTGATCGGATTTTAAATTCATTAGTTTATGCCAATTTT 527
QY 1060 TCAACGGATGCTGTATGATTAAGACCAACAGATGAAGCGCTGAATAGATTTT 1119
Db 528 CTCAATATTCAAGAAGAAGATGCTGCGGAAATAATTTTCTAAGCTCAATTAAGTCAG 587
QY 1120 CGGCTCGTTGTTATGCTGGTGGAG 1145
Db 588 TGGGTGATATTTTCTTTTAGCAG 613

RESULT 14
US-09-543-681A-3779
; Sequence 3779, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 3779
; LENGTH: 312
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-3779

Query Match 3.0%; Score 36; DB 4; Length 312;
Best Local Similarity 48.5%; Pred. No. 0.081;
Matches 99; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 917 AATCCAAAAGTATTGTTTCTGCGGCGCTATTTCCGCAATTCATCATGCCGCAACAG 976
Db 52 AATCCAAAAGCAATACTATCTTTACGGCTTTCTCCGCAATTCGTTAGCCCAATTA 111
QY 977 CGCAACTGATGCAATATATCGTCTCGGCGTCAACCACTATTGTGTCGATATTATTGTG 1036
Db 112 CCGGCTAGTTCGCCAGTTTCTAATACTTTGTTCTTTTACTCTCGAGTTTATTGCA 171
QY 1037 ATGATCGGTTAGCCACCCCTTGCTCAACGGATGCTCTATGGAATTAAGGACCAAGCAG 1096
Db 172 ATAATGCTCTACGTTGTTAGGCTTACACATGAAAAGTTGCAAAAAAACCAATGCT 231
QY 1097 ATGAAGCGCTGAATAAGATTTC 1120
Db 232 AAAAAAGTGTCAATCGTAATGTC 255

RESULT 15
US-09-655-908-17
; Sequence 17, Application US/09655908
; Patent No. 6645747
; GENERAL INFORMATION:
; APPLICANT: Hallahan, David L.
; TITLE OF INVENTION: cis-Prenyltransferases from Plants
; FILE REFERENCE: BC1019 US NA
; CURRENT APPLICATION NUMBER: US/09/655,908
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/155,046
; PRIOR FILING DATE: 1999-09-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 1028
; TYPE: DNA
; ORGANISM: Glycine max
US-09-655-908-17

Query Match 2.9%; Score 35.4; DB 4; Length 1028;
Best Local Similarity 54.1%; Pred. No. 0.26;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 954 CGCAATTCATCATGCCCAACAGCGCACTGATGCAATATATCGTCTCGGCGTCACCA 1013
Db 77 CCCTATTCCTCTCGTTAAACACCACTTCTCCCTCTTGTATTATCTCACTATTATCA 136
QY 1014 CTATTGTCGATATTATTGTGATGATCGGTTACGCCACCTTGTCTCAACGGATTGCTC 1073
Db 137 CTATCGTTATCGTTATCGTTGTTATCATCTCTTCCATCACCGTTCCTCCAAACACAGAGTCT 196
QY 1074 TATCGATTAAAGG 1086
Db 197 TATCGTCTCGAAG 209

Search completed: March 28, 2004, 18:24:27
Job time : 139.331 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 16:45:05 ; Search time 509.752 Seconds
(without alignments)
8764.284 Million cell updates/sec

Title: US-09-847-392-1

Perfect score: 1200
Sequence: 1 agaaataatgtggagatgcg.....atgtcggatgcggtgtaaac 1200

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1200	100.0	1200	9	US-09-927-395-1
2	1200	100.0	1200	9	US-09-847-392-1
C 3	446	37.2	1023	12	US-10-282-122A-20578
C 4	446	37.2	1023	15	US-10-369-493-47314
5	417	34.8	417	12	US-10-282-122A-6496
6	417	34.8	417	14	US-10-287-274-274
7	382.2	31.9	621	12	US-10-282-122A-39598
8	375	31.2	618	12	US-10-282-122A-23973
9	357.4	29.8	618	12	US-10-282-122A-20277
10	348.2	29.0	623	12	US-10-282-122A-36679
C 11	298.8	24.9	1017	12	US-10-282-122A-7563
C 12	297.2	24.8	1017	12	US-10-282-122A-39597
C 13	290.8	24.2	990	12	US-10-282-122A-20024
C 14	276.4	23.0	990	12	US-10-282-122A-23738
15	275	22.9	621	12	US-10-282-122A-41840

C 16	262.8	21.9	1026	12	US-10-282-122A-36678	Sequence 36678, A
C 17	259	21.6	259	12	US-10-282-122A-1269	Sequence 1269, Ap
C 18	207.4	17.3	263	14	US-10-287-274-107	Sequence 107, App
C 19	181.8	15.2	1014	12	US-10-282-122A-41838	Sequence 41838, A
C 20	165.8	13.8	975	15	US-10-369-493-24057	Sequence 24057, A
C 21	156.2	13.0	990	15	US-10-369-493-44919	Sequence 44919, A
C 22	149.8	12.5	618	12	US-10-282-122A-40780	Sequence 40780, A
C 23	144.8	12.1	1002	12	US-10-282-122A-32814	Sequence 32814, A
C 24	112.4	9.4	630	12	US-10-282-122A-30630	Sequence 30630, A
C 25	104	8.7	630	12	US-10-282-122A-33154	Sequence 33154, A
C 26	86.8	7.2	633	12	US-10-282-122A-31699	Sequence 31699, A
C 27	80	6.7	627	12	US-10-282-122A-14918	Sequence 14918, A
C 28	76	6.3	1017	12	US-10-282-122A-40779	Sequence 40779, A
C 29	70.8	5.9	978	15	US-10-369-493-35643	Sequence 35643, A
C 30	65.6	5.5	954	12	US-10-282-122A-31172	Sequence 31172, A
C 31	64.6	5.4	543	12	US-10-282-122A-8689	Sequence 8689, Ap
C 32	62.4	5.2	864	15	US-10-369-493-35283	Sequence 35283, A
C 33	62.4	5.2	864	15	US-10-369-493-38570	Sequence 38570, A
C 34	62.4	5.2	882	15	US-10-369-493-37975	Sequence 37975, A
C 35	62.4	5.2	882	15	US-10-369-493-38253	Sequence 38253, A
C 36	60.8	5.1	795	15	US-10-369-493-31514	Sequence 31514, A
C 37	58.8	4.9	633	12	US-10-282-122A-11642	Sequence 11642, A
C 38	58.2	4.9	792	15	US-10-369-493-45471	Sequence 45471, A
C 39	58.2	4.9	822	15	US-10-369-493-44326	Sequence 44326, A
C 40	53.6	4.5	1092	12	US-10-282-122A-15543	Sequence 15543, A
C 41	46	3.8	942	12	US-10-282-122A-22067	Sequence 22067, A
C 42	46	3.8	1830121	14	US-10-329-960-1	Sequence 1, Appli
C 43	46	3.8	1830121	15	US-10-329-960-1	Sequence 34253, A
C 44	45.6	3.8	810	15	US-10-369-493-34253	Sequence 119, App
C 45	43.2	3.6	536	16	US-10-338-110-119	

ALIGNMENTS

RESULT 1

US-09-927-395-1
; Sequence 1, Application US/09927395
; Patent No. US20020058314A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATREVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VINYAMTOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHAKOVA, IRINA LYOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO LO-HOMOSERINE AND METHOD
; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/927,395
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (557)..(1171)
US-09-927-395-1

Query Match 100.0%; Score 1200; DB 9; Length 1200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATAGCGTTTACGCCAC 60
DB 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATAGCGTTTACGCCAC 60

QY 61 GGACCGGCTGAACCTCCCTGCTSCAGAAATGCCCGCAGATCATCAACATAATCAATTAAG 120
DB 61 GGACCGGCTGAACCTCCCTGCTSCAGAAATGCCCGCAGATCATCAACATAATCAATTAAG 120
QY 121 CGATTAAATGATGCGGAGATGCGGATGCGGTAACAGCGACCGGACGTCCTGCCGCGA 180
DB 121 CGATTAAATGATGCGGAGATGCGGATGCGGTAACAGCGACCGGACGTCCTGCCGCGA 180
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCCGCAAT 240
DB 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCATAGCCAGTTCCGCAAT 240
QY 241 TTTACGTAGCTCTCAATAGCCCGGCGAGATGACTACACCCGTCATGCTGCTGTCG 300
DB 241 TTTACGTAGCTCTCAATAGCCCGGCGAGATGACTACACCCGTCATGCTGCTGTCG 300
QY 301 CGAAACGGAACAAAGCGACCGGAAATGTCATCCACACAGTAAATCTGCTTCATCAGC 360
DB 301 CGAAACGGAACAAAGCGACCGGAAATGTCATCCACACAGTAAATCTGCTTCATCAGC 360
QY 361 TGAGCCGAGAAATCAGTCAGCGTCCATGTTGTAAGAGCAGAAAGCGTTTCTCTGTT 420
DB 361 TGAGCCGAGAAATCAGTCAGCGTCCATGTTGTAAGAGCAGAAAGCGTTTCTCTGTT 420
QY 421 TCCAGTCTTTTGTGCTGAACATCGGTAATCTGCTCTTAAACACAGTAAATCGT 480
DB 421 TCCAGTCTTTTGTGCTGAACATCGGTAATCTGCTCTTAAACACAGTAAATCGT 480
QY 481 TTTTGTAGCGTCTGACACAAAGCGTGGTAAATGCTGCTCTTAAACACAGTAAATCGT 480
DB 481 TTTTGTAGCGTCTGACACAAAGCGTGGTAAATGCTGCTCTTAAACACAGTAAATCGT 480
QY 541 ACACCGGAGTTTATCATGACCTTAAATGCTGCTGCTTAAACACAGTAAATCGT 600
DB 541 ACACCGGAGTTTATCATGACCTTAAATGCTGCTGCTTAAACACAGTAAATCGT 600
QY 601 TTTAAGCTCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 601 TTTAAGCTCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 661 TTTAAGCTCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 TTTAAGCTCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 721 TGGTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGAATGCGTTGAAGTTGAA 780
DB 721 TGGTGGCGTGGGTTGGGACGCTATTTTCCGCTCAGTGAATGCGTTGAAGTTGAA 780
QY 781 GTGGCAGCGCGCTTACTGATTTGGCTGGGAAATCCAGCAGTGGCGCGCTGCTGTC 840
DB 781 GTGGCAGCGCGCTTACTGATTTGGCTGGGAAATCCAGCAGTGGCGCGCTGCTGTC 840
QY 841 AATTGACCTTAAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 841 AATTGACCTTAAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 901 TTTTGTGAATCTCAACAAATCCCAAGATGTTGTTTCTGGCGGCTATTTCCGCAAT 960
DB 901 TTTTGTGAATCTCAACAAATCCCAAGATGTTGTTTCTGGCGGCTATTTCCGCAAT 960
QY 961 CATCATGCGCAACAGCGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 961 CATCATGCGCAACAGCGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 1021 GGTGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
DB 1021 GGTGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
QY 1081 TAAAGGACCAAGCAGATGAAGCGCTGATTAAGATTTTCCGCTGCTGCTGCTGCTGCT 1140
DB 1081 TAAAGGACCAAGCAGATGAAGCGCTGATTAAGATTTTCCGCTGCTGCTGCTGCTGCT 1140

QY 1141 GGAGCGCTGTAGCATCGCGAGGCATGCTGTAATAATATGTCGATCGCGGTAAC 1200
DB 1141 GGAGCGCTGTAGCATCGCGAGGCATGCTGTAATAATATGTCGATCGCGGTAAC 1200
RESULT 2
US-09-847-392-1
; Sequence 1, Application US/09847392
; Patent No. US20020102670A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENTAMIOVICH
; APPLICANT: BELAROVA, ALL VALENTINOVNA
; APPLICANT: TOKHMAKOVA, IRINA LVOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/847,392
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 09/196,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (557) ..(1171)
US-09-847-392-1

Query Match 100.0%; Score 1200; DB 9; Length 1200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAATAATATGGAGATCGACCGCCATCGAATGTCAGTATATAGCGTTTACGCCAC 60
DB 1 AGAATAATATGGAGATCGACCGCCATCGAATGTCAGTATATAGCGTTTACGCCAC 60
QY 61 GGACCGGCTGAACCTCTGCTGCGAGATGCCCGCAGATCATCAATATTAAG 120
DB 61 GGACCGGCTGAACCTCTGCTGCGAGATGCCCGCAGATCATCAATATTAAG 120
QY 121 CGATTAAACATGCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCTGCCCGCA 180
DB 121 CGATTAAACATGCCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCTGCCCGCA 180
QY 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCATAGGCCAGTTCCGCATAT 240
DB 181 TGGTCGATGATTAAGACATCAAAACCCCAATGGAACAGGTCATAGGCCAGTTCCGCATAT 240
QY 241 TTTAGTGTAGCTCTCAATAGCGCCCGGCGAGATGACTACACCGGCTCATGCTGCTGCG 300
DB 241 TTTAGTGTAGCTCTCAATAGCGCCCGGCGAGATGACTACACCGGCTCATGCTGCTGCG 300
QY 301 CGAAAACGGAACAAAGCGCACCGGAATGTCTCCACACAGTAAATCTGCTTCATCAGC 360
DB 301 CGAAAACGGAACAAAGCGCACCGGAATGTCTCCACACAGTAAATCTGCTTCATCAGC 360
QY 361 TGACGCCGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 TGACGCCGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 TCCAGTCTTTTGTGCTGTAACATCGGTAATCTGCTCTTAAACACAGTAAATCGT 480
DB 421 TCCAGTCTTTTGTGCTGTAACATCGGTAATCTGCTCTTAAACACAGTAAATCGT 480
QY 481 TTTTGTAGCGTCTGACACAAAGCGTGGTAAATGCTGCTCTTAAACACAGTAAATCGT 540

Db 481 TTTTATAGCGTCTGACACACGCTGCGACGATGAGTATTTGGCACAAAATAGAC 540
Qy 541 ACACCGGAGTTTCATGACCTTAAATGGTGGTTTCCCTACCTGCTGACATGATCAT 600
Db 541 ACACCGGAGTTTCATGACCTTAAATGGTGGTTTCCCTACCTGCTGACATGATCAT 600
Qy 601 TTTAAGCTGTCCGAGGCTCTGTGGCAATCAACTATGACACCTGCTCAACACGG 660
Db 601 TTTAAGCTGTCCGAGGCTCTGTGGCAATCAACTATGACACCTGCTCAACACGG 660
Qy 661 TTTATCCGCGGCTGTGGCTTATTTGGGCTTCCAGCCGCTGCGATTCATATGTC 720
Db 661 TTTATCCGCGGCTGTGGCTTATTTGGGCTTCCAGCCGCTGCGATTCATATGTC 720
Qy 721 TGGTGGCTGTGGGCTTATTTCCGCTCAGTGAATGGTTGAAGTTGAA 780
Db 721 TGGTGGCTGTGGGCTTATTTCCGCTCAGTGAATGGTTGAAGTTGAA 780
Qy 781 GTGGGAGCGGCTTACTTATTTGGCTGGAATCCAGCAGTGGCGCGCTGCTGC 840
Db 781 GTGGGAGCGGCTTACTTATTTGGCTGGAATCCAGCAGTGGCGCGCTGCTGC 840
Qy 841 AATTGACCTTAAATCGCTGCTCTACTCAATCGCGTGCACATTTGTCAGCGCGCAGT 900
Db 841 AATTGACCTTAAATCGCTGCTCTACTCAATCGCGTGCACATTTGTCAGCGCGCAGT 900
Qy 901 TTTTGTGAATCTCAACCAATCCCAAAAGTATTTGTTCTGCGCGCTATTTCCGCAATT 960
Db 901 TTTTGTGAATCTCAACCAATCCCAAAAGTATTTGTTCTGCGCGCTATTTCCGCAATT 960
Qy 961 CATCATCGCGCAACAGCGCAACTGATGAGTATATCGTCTCGCGCTCACCACTATTGT 1020
Db 961 CATCATCGCGCAACAGCGCAACTGATGAGTATATCGTCTCGCGCTCACCACTATTGT 1020
Qy 1021 GGTGATATATTTGATGATCGTTAGCGCCACCTTCTCTCAACGGAATGCTATGAT 1080
Db 1021 GGTGATATATTTGATGATCGTTAGCGCCACCTTCTCTCAACGGAATGCTATGAT 1080
Qy 1081 TAAAGACCAAGCAGATCAAGCGCTGAATGAATTTTCGCTGCTTTTATGCTGGT 1140
Db 1081 TAAAGACCAAGCAGATCAAGCGCTGAATGAATTTTCGCTGCTTTTATGCTGGT 1140
Qy 1141 GGGAGCGCTTTAGCATCGCGGAGGATCGCTGGAATAATGTCGATGCGCGGTAAC 1200
Db 1141 GGGAGCGCTTTAGCATCGCGGAGGATCGCTGGAATAATGTCGATGCGCGGTAAC 1200

RESULT 3

US-10-282-122A-20578/c
; Sequence 20578, Application US/10282122A
; Publication No. US20040029129A1

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20578
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-282-122A-20578

Query Match 37.2%; Score 446; DB 12; Length 1023;

Best Local Similarity 100.0%; Pred. No. 2.4e-144; Indels 0; Gaps 0;
Matches 446; Conservative 0; Mismatches 0;

Qy 1 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
Db 446 AGAAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 387
Qy 61 GGACCGGCTGAACCTCTGCTGCAGATGCGCCAGATCAACACATATCAATTAAG 120
Db 386 GGACCGGCTGAACCTCTGCTGCAGATGCGCCAGATCAACACATATCAATTAAG 327
Qy 121 CGATTAAACATCCCGAGATGCGGATCGGCTAACAGGCGACCGAAGCTCCCTCCCGCGA 180
Db 326 CGATTAAACATCCCGAGATGCGGATCGGCTAACAGGCGACCGAAGCTCCCTCCCGCGA 267
Qy 181 TGGTGCATGATTAAGACATCAACCCCAATGGAACAGGTATAGCCAGTCCGCGATAT 240
Db 266 TGGTGCATGATTAAGACATCAACCCCAATGGAACAGGTATAGCCAGTCCGCGATAT 207
Qy 241 TTTAGTACCTCTCAATACGCGCGGCGAGATGACTACACCGGTCATGCTGTGCG 300
Db 206 TTTAGTACCTCTCAATACGCGCGGCGAGATGACTACACCGGTCATGCTGTGCG 147
Qy 301 CGAAAAACGGAACAGCGCACCGGAATGTCTATCCACACAGTAACTCTGCTTATCAGC 360
Db 146 CGAAAAACGGAACAGCGCACCGGAATGTCTATCCACACAGTAACTCTGCTTATCAGC 87
Qy 361 TGACGCCAGAAATCAGTCAGCGGTCCCATGTTAAAGCAGCAACGCGTTTCTCTTGT 420
Db 86 TGACGCCAGAAATCAGTCAGCGGTCCCATGTTAAAGCAGCAACGCGTTTCTCTTGT 27
Qy 421 TCCAGTCTTTTGTGCTGTAACAT 446
Db 26 TCCAGTCTTTTGTGCTGTAACAT 1

RESULT 4

US-10-369-493-47314/c

; Sequence 47314, Application US/10369493

; Publication No. US20030233675A1

GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

```
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 47314
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-369-493-47314

Query Match      37.2%; Score 446; DB 15; Length 1023;
Best Local Similarity 100.0%; Pred. No. 2.4e-144;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGAGATCGCACCGCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
DB 446 AGAAATAATGTGAGATCGCACCGCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 387
QY 61 GGACCGGGTGAACCTCTGTGCGCCAGAAATGCCGCCAGATCATCAATCAATTAAG 120
DB 386 GGACCGGGTGAACCTCTGTGCGCCAGAAATGCCGCCAGATCATCAATCAATTAAG 327
QY 121 CGATTACATGCCCGAGATCGGATCGCTACAGCGACCGACGTCCTGCGCCGA 180
DB 326 CGATTACATGCCCGAGATCGGATCGCTACAGCGACCGACGTCCTGCGCCGA 267
QY 181 TGTCGATGATTAAAGCATCAACCCCAATGAACAGTATAGGCCAGTTCGCGCATAT 240
DB 266 TGTCGATGATTAAAGCATCAACCCCAATGAACAGTATAGGCCAGTTCGCGCATAT 207
QY 241 TTTACGTAGTCTCAATAGCCCGCGGAGATGACTACACCGGTCTATGTGTGTGG 300
DB 206 TTTACGTAGTCTCAATAGCCCGCGGAGATGACTACACCGGTCTATGTGTGTGG 147
QY 301 CGAAACGGCAAAAGCGCACCGGAATGTATCCACACAGCAAAATCTGTCTTCATCAGC 360
DB 146 CGAAACGGCAAAAGCGCACCGGAATGTATCCACACAGCAAAATCTGTCTTCATCAGC 87
QY 361 TGACGCCAGAAATCAGTCAGCGTCCATGTTAAAGAGCAAGCGTTTCTCTGTT 420
DB 86 TGACGCCAGAAATCAGTCAGCGTCCATGTTAAAGAGCAAGCGTTTCTCTGTT 27
QY 421 TCCAGTCTTTTGTGCTGTAACAT 446
DB 26 TCCAGTCTTTTGTGCTGTAACAT 1

RESULT 5
US-10-282-122A-6496
; Sequence 6496, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 47314
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-369-493-47314

Query Match      34.8%; Score 417; DB 12; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.8e-134;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 GTGATTGGTTTGAAGTGTTCGAAGTGGGCGGCGCTTACTTGTGTTGGTGGGAATC 817
DB 1 GTGATTGGTTTGAAGTGTTCGAAGTGGGCGGCGCTTACTTGTGTTGGTGGGAATC 60
QY 818 CAGCAGTGGCGCGCGCTGGTGGCAATTAATCGTGGCTCTACTCAATCCGCT 877
DB 61 CAGCAGTGGCGCGCGCTGGTGGCAATTAATCGTGGCTCTACTCAATCCGCT 120
QY 878 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCATCCCAAAGTATTGTGTT 937
DB 121 CGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCATCCCAAAGTATTGTGTT 180
QY 938 CTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCGCAACTGTATGATATC 997
DB 181 CTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCGCAACTGTATGATATC 240
QY 998 GTGCTCGGCGTCACCACTATTGTGTCGATATTATTTGATGATCGGTACGCCACCTT 1057
DB 241 GTGCTCGGCGTCACCACTATTGTGTCGATATTATTTGATGATCGGTACGCCACCTT 300
QY 1058 GCTCAACGAGTTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 1117
DB 301 GCTCAACGAGTTGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATAAGATT 360
QY 1118 TTCGCTCGTTGTTTATCTGCTGGGAGCGCTGTAGCATCGGCGAGCATCGGTGA 1174
DB 361 TTCGCTCGTTGTTTATCTGCTGGGAGCGCTGTAGCATCGGCGAGCATCGGTGA 417

RESULT 6
US-10-287-274-274
; Sequence 274, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELITRA.008DVI
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
```

854	Qy	TCGTCGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGATTTTTGTGAATCTC	913
301	Db	ACTCTCGCCAGACGCAATCGCGGGTCGGCTGTTCAAACGGGCGGATTTGTCAATCTA	360
914	Qy	ACCAATCCCAAAAGTATGTGTTCTTCGGCGGGCTATTTCGCGAAATTCATCATCCGCGAA	973
361	Db	ACCAATCCCAAAAGTATGTGTTCTTCGGCGGGCTGTTCGCGAAATTCATCATCCGCGAG	420
974	Qy	CAGCGCGCAACTGATGTCAGTATATCGTGTTCGGCGCTCACCACCTATTGTGGTCGATATTATT	1033
421	Db	CAACGCGCAACTGGCGCAGTACCTCATTTCTCGCGCTCACCACGATTTGGTGGATATGATT	480
1034	Qy	GTGATGATCGGTTACGGCCACCCCTTGCTCAACGGATTGCTCTATGGAATTAAGGACCAAG	1093
481	Db	GTGATGACCGGTTACGCCACGCTGGCGCAGCGCATTCGCGCTCGATTAAAGGACCAAG	540
1094	Qy	CAGATGAAGCGCTGAATTAAGATTTTCGGCTCGTTGTTTATGCTGTGGGAGCGCTCTTA	1153
541	Db	CAGATGAAGCGCTGAATTAAGCGTTTGGTTCGTTGTTTATGCTGTAGGCGCTCCTG	600
1154	Qy	GCATCGCGAGGCATGCGTGA	1174
601	Db	GCGTCCGCAAGACACGCGTGA	621

RESULT 8

```
US-10-282-122A-23973
; Sequence 23973, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23973
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23973

Query Match          31.2%; Score 375; DB 12; Length 618;
Best Local Similarity 77.6%; Pred. No. 1.1e-119;
Matches 479; Conservative 0; Mismatches 135; Indels 3; Gaps 2;

Qy 557 ATGACCTTAGAATGGTGGTTGGCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA 616
Db 1 ATGACCAATTGAGTGGTGGTTGGCTACCTGCTGACATCAATTAATTTCTCAGCTGTGCGCG 60

Qy 617 GGCTCTGGTGCATCAACACTATGACACCTCGCTCAACACCGGTTA--TCGGGCGCGTG 674
Db 61 GGTTCCGAGCGAATTAATACCATGACCACTCCATTAAACAGGATACCGGCGCGCG 120

Qy 675 CGGCTATTGCTGGGCTTAGACCGGACTGGCGATTCATATTGCTGCTGGCG-TGGG 733
Db 121 GCGTCGATTCCGGCTTGCAGACCGGCGTGGTTATTATATCATCTGCTGGTGGCGTTGGT 180

Qy 734 GTTGGGACGCTATTATTTCCCGCTCAGTGGATTGGCTTTGAAGTGTGAAGTGGCGCGCG 793
Db 181 CTCGGCACTCTCTCTCCCGCTGGTGGCTTTGAGTGTGCTGAATGGCGCGCGCC 240

Qy 794 GCTTACTTGAATTTGGTGGGAATCCAGCAGTGGCGCGCGCTGGTGGCAATTGACCTTAAA 853
Db 241 GCCTACCTGATCTGGCTGGGTATCCAGCAATGGCGGCGCGCGCGCGATGACCTGAAT 300
```

```
Qy 854 TCGCTGGCTCTTACTCAATCGCGTCGACATTGTTTCAGCGCGCGAGTTTGTGAATCTC 913
Db 301 ACGCTGGCGAAGGCGCGAGCGCGGCAAACTGTTCCAGCGCGCGTGTTCGTAATCTC 360

Qy 914 ACCAATCCAAAAGTATTGTTTCTGGCGCGCTATTTCGGCAATTCATCATGCCGCA 973
Db 361 ACCAATCCAAAAGCATTTGTTTCTGGCGCGCTGTTTCCGCGAGTTTATCTCTGCGCAC 420

Qy 974 CAGCGCGCAACTGATCGAGTATATCGTGTGGCGCTCACCACCTATTGTGGTGGATATT 1033
Db 421 CAGCGCGCAGGTGATGCAATATCTGTTCTGGCGCTCACCACCTCGTCGTCGATACATT 480

Qy 1034 GTGATGATCGGTATAGCCACCCCTTGTCTCAACGAGTATCTCTATGATTAAGACCAAG 1093
Db 481 GTGATGATGGCTATGCGACCCCTGCGCAGCGCATTTCCGATGATGATCAAAAGGCCAAAG 540

Qy 1094 CAGATGAAGGCGCTGAATAAGATTTCGGCTCGTGTGTTTATGCTGGTGGAGCGCTGTTA 1153
Db 541 CAGATGAAGGCGCTGMAACAAAGTCTTTGGTTCATCTGTTTATGCTGGTGGCGGCGCTC 600

Qy 1154 GCATCGCGCAGGCGATGC 1170
Db 601 GCCTCGCGCGCGCCATGC 617

RESULT 9
US-10-282-122A-20027
; Sequence 20027, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20027
; LENGTH: 618
; TYPE: DNA
```

ORGANISM: Enterobacter cloacae
US-10-282-122A-20027

Query Match 29.8%; Score 357.4; DB 12; Length 618;
Best Local Similarity 75.9%; Pred. No. 1.6e-113;
Matches 468; Conservative 0; Mismatches 146; Indels 3; Gaps 2;

QY 557 ATGACCTTAGAATGGTGGTTCCTACCTGCTGACATCGATCATTTTAAACGCTGTGGCCA 616
DB 1 ATGACCTTCGATGGTGGTTCCTACCTGCTGACATCGATCATTTTAAACGCTGTGGCCA 60
QY 617 GCGTCTGGTGCATCAACATGATGACCACTGCTGCTCAACACGGTTA--TCCGGCCGGTG 674
DB 61 GCGTCTGGGGGCTATTAAACACCATGACCACTCCATCAATCAGCGCTACCGCGGTGGGG 120
QY 675 GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGCTGCTGGTGGGTGG 733
DB 121 GCAATGATTCGCGTTTCAGACCGGCTGGGTATTCATATTGCTGCTGGGTGGGTGG 180
QY 734 GTTGGAGCGCTATTTCCCGCTCAGTGAATCGTTTGAAGTGTGAAGTGGGAGCGCG 793
DB 181 CTGGGGACCTCTTCTCCCGCTCCGCTGCTGGCCCTTGAAGTGTGAAGTGGGAGCG 240
QY 794 GCTTACTTGAATGGTGGGATCCAGACGAGTGGCGCGCTGGTGAATGACCTTAA 853
DB 241 GCGATCTGATTTGGCTCGGATCCAGACGAGTGGCGCGGCGGCTCCATCAACCTGA 300
QY 854 TCCTGGGCTCTACTCAATCCGCTGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 913
DB 301 ACCTGGCTCTGACGCAAAACCGGGCCATCTGTTTAAAGTGGGTATTCGTCACCTG 360
QY 914 ACATATCCAAAGATTGTGTTTCTGGCGGCTATTTCCGGAAATTCATCAGCGCGCA 973
DB 361 ACCAACCGGAGAGATCGTGTCTCCCGCGCTGTTCGCGCAGTTTATCGTTCGCGAT 420
QY 974 CAGCGGCAACTGATGCAATATGCTGCTCGGCGTCAACACTATTGCTGCGATATT 1033
DB 421 CAGCCTCAGGTGATGACAGTACGTGCTGCTGGCGCGACCATATTATCGTGCATATC 480
QY 1034 GTGATGCTGGTTCAGCACCTGCTCAACGATGCTCTATGATTAAGACCAAG 1093
DB 481 GTGATGATGTTAGCGACGCTGGCGAGCAATTCGCGCTGGATTAAGGCGCTTAAG 540
QY 1094 CAGATGAAGGCGCTCAATAAGATTTTCGGCTGCTGTTTATGCTGGGAGCGCTG 1153
DB 541 CAGATGAAGGCGCTCAATAAGATTTTCGGCTGCTGTTTATGCTGGTGGTGGCTG 600
QY 1154 GCATCGGCGAGCATGC 1170
DB 601 GCGTCAGCGCGTCACGC 617

RESULT 10

US-10-282-122A-36679
; Sequence 36679; Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36679
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-36679
Query Match 29.0%; Score 348.2; DB 12; Length 623;
Best Local Similarity 76.6%; Pred. No. 2.6e-110;
Matches 479; Conservative 0; Mismatches 138; Indels 8; Gaps 4;
QY 557 ATGACCTTAGAATGGTGGTTCCTACCTGCTGACATCGATCATTTTAAACGCTGTGGCCA 616
DB 1 ATGACCTTGAATGGTGGTTCCTACCTGCTGACATCGATCATTTTAAACGCTGTGGCCA 60
QY 617 GCGTCTGGTGCATCAACATGATGACCACTGCTCAACACCGTTTCCGGCG--GTG 674
DB 61 GCGTTCAGCGCGCATCAATACATCAACGAGCTCTATCAACCATGATATCGTGGCGAGCG 120
QY 675 GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGCTGCTGGTGGCGTGGG 734
DB 121 GCTTCTATCGCGGACTCCAGACCGGACTGGGATACATATCTGCTGGTGGCGTGGCA 180
QY 735 TTGGG-ACGCTATTTTCCGCTCAGTGAATGCGTTTGAAGTGTGAAGTGGGAGCGCG 793
DB 181 CTGGGTACGCTCTTTTCGCGCTCGCTCTCTCTTTTGAATTCGAAATGGGCTGGCGCG 240
QY 794 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTCAATTTGACCTTAA 853
DB 241 GCTTATCTTATCTGGCTGGGTATCCAGCATGGCG--CGCAGCGCTATCGATCTGCAT 297
QY 854 TCCTGGGCTCTACTCAATCGGTGACATTTGTTCCAGCGCGCAGTTTGTGGAATCTC 913
DB 298 ACTCTCGCCAGACGCAATCGCGGGTTCGGCTGTTTCAACACGGCGGATATTTGTCAATCTA 357
QY 914 ACCAATCCAAAAGTATTGTTTCTGGGGCGCTATTTCCGCAATTCATCATCGCGCAA 973
DB 358 ACCAATCCAAAAGTATTGTTTCTGGCGCCTGTTTCCGCAATTCATCATCGCGCAG 417
QY 974 CAGCGCAACTGATGACGATATATCGTCTGGCGCTCACCACCTATTGCTGCGATATT 1033
DB 418 CAACCGCAACTGGCGCAGTACCTCAITTCGGCGTCAACACGATTCGTGGTGAATGAT 477
QY 1034 GTGATGATCGGTACGCGACCCCTGCTCACGG--ATTGCTCTATGATTAAGGACCAA 1091
DB 478 GTGATGACCGGTACGCGACACCTGGCGCTAGCGCATTTGCGCGCTGGATTAAGGACCAA 537
QY 1092 AGCAGATGAAGGCGCTGAATAAGATTTTCGGCTGTTGTTTATGCTGGTGGAGCGCT 1151
DB 538 AGCAGATGAAGGCGCTGAATAAGGCGTTGTTGTTGTTTATGCTGGTGGAGCGCTCC 597

QY 1 AGAATAATGTGGAGATCGACCGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 60
Db 446 AGGAACAGGGTGGCTATCGCGCGCCCATTCAGTGGCCAGGATATAAGCGTTAGGCCAC 387
QY 61 GGACCGGCTGAACCTCTGCTGCGAGATCGCCCGAGATCATCAATATCAATTAAG 120
Db 386 GACCCGGCTCAATTCCTGCTGCGAGAACCGCGTAACTCTCAACATATCGTTAAAA 327
QY 121 CGATTAACATGCCGAGATCGCGATCGGCTAACAGGCGACCGGAACGTCCTGCGCGCGA 180
Db 326 TGATCGACATGACCCGATCGGGTCCGATAACATCGCCCGGAGCGCCCTTGACCGAG 267
QY 181 TGGTCGATGATTAAAGACATCAACCCCAATGGAACAGGTCATAGGCGAGTTCGGCATAT 240
Db 266 TGGTCGATGATGAAGATATCAAAACCGAGATGAAAGAGTTCATACGCCAGTTCAGCGTAT 207
QY 241 TTTAGCTAGCTCTCAATACGCGCCCGGCGAGATGACTACACCGCGTCATGCTGTGCG 300
Db 206 TTTACATRACTCTCAATGCGCGCAGGACAAATGACGATCGTCCGATCGTTGGAGTCATTC 147
QY 301 CGAAACGAGCAAAAGCGCACCGGAATGTCTATCCACACAGTAAACTCTGCTTCAATACGC 360
Db 146 CGAAACGAGCAAAAGCGCACCGGAATGTCTATCCAGCGCGATAAACTCCGCTCTTCCCGC 87
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCTCCATGTGTAAGCAGCAAAACGGTTTCTCTTGT 420
Db 86 TGACGCCAGAAATCGTTCAGCGGCGCCCATAGCAAAAGCGGCAACGATTTCTCTTGT 27
QY 421 TCCAGTCTTTTGTGCTGTAACAT 446
Db 26 TCCAGTCTTTTGTGCTGTAACAT 1

RESULT 13
US-10-282-122A-20024/c
; Sequence 20024, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 20024
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-10-282-122A-20024
Query Match 24.2%; Score 290.8; DB 12; Length 990;
Best Local Similarity 78.3%; Pred. No. 4e-90;
Matches 349; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 1 AGAATAATGTGGAGATCGCACCGCCCATCGAATGTGCCAGTATATAGCGTTACGCCAC 60
Db 446 AAAAAACAGCTCGAAATGGCGCGCCCATCGAGTGGCGAAGATATACCGCTTGGCGCAG 387
QY 61 GGACCGGCTGAACCTCTGCTGCGAGATCGCCCGAGATCATCAATATCAATTAAG 120
Db 386 GGGCCGGGCTGAACGTCCTGCTGCGAGACGGGCAAAATCGTCGACGTAATCGCTGAAG 327
QY 121 CGATTAAATATCGCCGAGATCGGATCGGCTAACAGGCGACCGGAACGTCCTGCGCGCGA 180
Db 326 CGATCGACGTCCCGACCGGTGCTGTCGGCAGCATACGTCGGGACAGCCCTGTCCACGA 267
QY 181 TGGTCGATGATTAGACATCAAAACCCCAATGGAACAGGTCATAGGCGAGTTCGGCATAT 240
Db 266 TGGTCGATGATCAGGACATCGAACCCCATATGACACCGAGTTCATAGGCCAGCTCGCGTAT 207
QY 241 TTTAGCTAGCTCTCAATACGCGCCCGGCGAGATGACTACACCGCGTCATGCTGTGCG 300
Db 206 TTTAGCTAGCTCTCAATACGCGCGGCGCAGACCAATACCCCGGTCATTTTTCATTT 147
QY 301 CGAAACGAGCAAAAGCGCACCGGAATGTCTATCCACACAGTAAACTCTGCTTCAATACGC 360
Db 146 CGAAACGAGCAAAAGCGCACCGGATGCTCCCGGATGCTCCCGCGCGCGCGTCTCTCTGTT 87
QY 361 TGACGCCAGAAATCAGTCAGCGGTCCTCCATGTGTAAGCAGCAAAACGGTTTCTCTTGT 420
Db 86 TGGCGCCAGAAATCGTTCAGCGGCGCCCATAGCAAAAGCGGCAACGATTTCTCTTGT 27
QY 421 TCCAGTCTTTTGTGCTGTAACAT 446
Db 26 TCCAGTCTTTTGTGCTGTAACAT 1

RESULT 14
US-10-282-122A-23738/c
; Sequence 23738, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727


```
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 23738
/ LENGTH: 990
/ TYPE: DNA
/ ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23738

Query Match      23.0%; Score 276.4; DB 12; Length 990;
Best Local Similarity 76.2%; Pred. No. 4.3e-85;
Matches 340; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 1 AGAATAATGTGAGATCGACCGCCATCGAATGTGCCAGTATATACGTTTACGCCAC 60
DB 446 AGAAAGCGTGGCAATCGCTCCGCCATAGAAATGGCAAGATAAACCGTTTTCGCCAG 387
QY 61 GGACCGGCTGAAACCTTCCTGCTGCCAGATGCGGCCAGATCATCAACATAATCAATAAG 120
DB 386 TGTCAGGACCACTGCTGCTGCCAGCGCGCCAGGTATCAACATAGCTCGCTGAG 327
QY 121 CGATTACATGCCCGAGATCGGATCGGCTAACAGCGACCGAAACGTCCTTGCCTGCCGA 180
DB 326 TTAACACAGTGTCCCGGCTGCTGATCTGAGACGCGCGCGCGCGTCCCTGCGCCGCA 267
QY 181 TGCTCATGATTAAAGACATCAACACCCCAATGGAACAGGTATAGGCCAGTTCGCGCAT 240
DB 266 TGCTCATGATCATCGTCAAGCGCGAGTGAAGAGATCGTAGGCACCTTCAGCATAC 207
QY 241 TTACGTAGTCTCAATACGCCCGCGCGAGATGATACACCCCGGTATGCTGTGTGCG 300
DB 206 TTCACGTAGTCTCAATAGCGCGCGCGGCAATCAACACCGACGATCGTTATGTTGCGCG 147
QY 301 CGAAAGCGACAAAGCGCACCGGAATGTCATCCACACAGTAAACTCTGCTTCATCAGC 360
DB 146 CAAAGCGGACAAAGCGTACCGGTATCATCTCGACGCTTAAATTCAGCTTCTTCGGC 87
QY 361 TGACGCCAGAAATCAGTCAGCGTCCCATGTTAAAGCAGCAACCGTTTCTCTTGT 420
DB 86 TGTGCGCAGAAATCCGTGAGCGGCCCATGCTAAAGCGCGCAACCGCTTCTCGTGT 27
QY 421 TCCAGCTCTTTTGTGCTGGAACAT 446
DB 26 TCCAGTCTCTTTTGTGCGCAACAT 1

RESULT 15
US-10-282-122A-41840
/ Sequence 41840, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
```

```
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 41840
/ LENGTH: 621
/ TYPE: DNA
/ ORGANISM: Yersinia pestis
US-10-282-122A-41840

Query Match      22.9%; Score 275; DB 12; Length 621;
Best Local Similarity 67.3%; Pred. No. 9.9e-85;
Matches 418; Conservative 0; Mismatches 200; Indels 3; Gaps 2;

QY 557 ATGACCTTAGATGGTGGTTTGGCTACTGCTGACATCATTTTAAACGCTGCCCA 616
DB 1 ATGACCTTAGATGGTGGTTTAACTTATCTGCTGACAAATGATCCTCAGTCTTTCTCCC 60
QY 617 GCCTCTGTGTGAATCAACACTATGACCACCTCGCTCAACCACCGTTATC--CGGCCGGTG 674
DB 61 GCCTCGGGGCCATTAAACAATGAGCACTGCCATAGCCATGCTACCGTGGGGTTGTT 120
QY 675 GCGTCTATGTGGGCTTCAGACCGGACTGGCGATTATATTGCTGTTGGGT--GGG 733
DB 121 GCTTCCATGGCGGGTTACAACTGGGCTGCGGTGATATCGTCTTGGTGGGTGGGT 180
QY 734 GTTGGAGAGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGCGAGCGCG 793
DB 181 CTGTGTGCTTGGTCTCTCAATCTTACTGCAATTTGAATATTAAAGTGGCTGGGTGCG 240
QY 794 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTGTGCAATTGACCTAAA 853
DB 241 GCTTACTTAAATTTGGCTTGGTATTCAACAATGGCGGTGACGCCGGGTGCTTGACCTGAT 300
QY 854 TCGCTGGGCTCTACTCAATCCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTC 913
DB 301 GCATCGGCAACAGTATGCCAGGCTAAGCTGTTTAAAGTGGGTGTTTGTAACTG 360
QY 914 ACCAATCCAAAAGTATTGTTTCTGGCGCGCTATTTCGGCAATTCATCATGCCGCAA 973
DB 361 ACCAATCCAAAAGCATTTGTTTCTGGCGCGTATTTCACCAATTTGACTGCCACAG 420
QY 974 CAGCGCGACTGATGCTGATATATCGTCTGGCGCTCACCATTATTTGTTGCTGATTATT 1033
DB 421 CAACCGCAGGTGGCAAGTATTGTTTGGGAGCAGCCAGTGTGATGTCGATATTATC 480
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:07:39 ; Search time 3810.91 Seconds
(without alignments)
9403.168 Million cell updates/sec

Title: US-09-847-392-1

Perfect score: 1200

Sequence: 1 agaaataatgagagatgc.....atgcggatcgccgtaaac 1200

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_nam.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_rtd.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	634.6	52.9	769	28	AQ858095
2	371	30.9	473	28	AQ859250
3	339	28.2	687	28	AY080100
4	218.2	18.2	815	29	CNS011BR

5	120.6	10.0	954	28	AF114214
6	85.4	7.1	1095	28	BZ548406
7	55.8	4.7	1199	28	BZ566001
8	53	4.4	231	14	CB042356
9	44.8	3.7	667	28	AQ991154
10	41.4	3.4	922	29	CNS014YP
11	41	3.4	849	13	EX462111
12	40.6	3.4	1039	29	CNS016GW
13	39.8	3.3	475	28	BH209707
14	39.8	3.3	990	29	AY401573
15	39.4	3.3	712	13	EX416727
16	39.4	3.3	885	13	EX425603
17	39.2	3.3	1101	29	CNS0021D
18	39	3.2	588	14	CB281433
19	38.8	3.2	1146	13	EX446722
20	38.8	3.2	685	13	BY724027
21	38.8	3.2	802	29	EX167575
22	38.8	3.2	1101	29	CNS00B90
23	38.6	3.2	642	9	AL628057
24	38	3.2	510	12	BG664436
25	38	3.2	1201	13	EX376097
26	37.8	3.1	1101	29	CNS017QV
27	37.8	3.1	1201	9	AL567139
28	37.6	3.1	798	10	BE659545
29	37.4	3.1	586	14	CD651858
30	37.2	3.1	1159	29	CNS015XR
31	37.2	3.1	1201	13	EX381961
32	37.2	3.1	1201	13	EX394709
33	37	3.1	264	10	AW347581
34	37	3.1	308	10	AW266920
35	37	3.1	437	12	BG688576
36	37	3.1	450	10	AW482031
37	37	3.1	495	10	BF060721
38	37	3.1	515	10	BE663253
39	37	3.1	522	12	BM087647
40	37	3.1	539	12	BM253190
41	37	3.1	573	10	BE665319
42	37	3.1	573	12	BI773985
43	37	3.1	585	12	BI774021
44	37	3.1	599	14	CB466945
45	37	3.1	603	9	AV596505

ALIGNMENTS

RESULT 1	AQ858095	nbeb001114f	CUGI Rice BAC Library (EcoRI)	Oryza sativa (japonica cultivar-group)	genomic clone nbeb001114f, genomic survey sequence.
LOCUS	AQ858095	769 bp	DNA	linear	GSS 03-NOV-1999
DEFINITION	AQ858095	1	GI:6208552		
ACCESSION	AQ858095				
VERSION	AQ858095.1				
KEYWORDS	GSS.				
SOURCE	Oryza sativa (japonica cultivar-group)				
ORGANISM	Oryza sativa (japonica cultivar-group)				
REFERENCE	1	(bases 1 to 769)			
AUTHORS	Wing, R.A. and Dean, R.A.				
TITLE	A BAC End Sequencing Framework to Sequence the Rice Genome				
JOURNAL	Unpublished (1998)				
COMMENT	Contact: Wing RA Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4233 Email: rwing@clemson.edu Seq primer: TAATACGACTCACTATAGG Class: BAC ends				

High quality sequence start: 42
 High quality sequence stop: 454.
 Location/Qualifiers
 1. .769

FEATURES

source

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/strain="Japonica"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone="nbe001114f"

/tissue_type="Leaf"

/lab_host="E. coli DH10B"

/note="lib="CUGI Rice BAC Library (EcoRI)"

/note="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Query Match 52.9%; Score 634.6; DB 28; Length 769;

Best Local Similarity 99.1%; Pred. No. 9.4e-183;

Matches 659; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1 AGAATAATGTGGAGATCGCAGCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 60
 DB 102 AGAATAATGTGGAGATCGCAGCGCCCATCGAATGTGCCAGTATATAGCGTTTACGCCAC 161
 QY 61 GGACCGGCTGAACCTCTGTCGAGATGCGGCCAGATCATCAATATCATTAAG 120
 DB 162 GGACCGGCTGAACCTCTGTCGAGATGCGGCCAGATCATCAATATCATTAAG 221
 QY 121 CGATTAAATGCGCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCCTGCCCGCA 180
 DB 222 CGATTAAATGCGCGAGATCGGATCGGCTAACAGCGACCGGAACGTCCCTGCCCGCA 281
 QY 181 TGGTCGATGATTAAAGCATCAACCCCAATGGAACAGGTATAGGCGAGTTCCGCATAT 240
 DB 282 TGGTCGATGATTAAAGCATCAACCCCAATGGAACAGGTATAGGCGAGTTCCGCATAT 341
 QY 241 TTATACCTAGCTCTCAATAGCCCGGGGAGATGACTACCCCGTTCATGTCGTGGCG 300
 DB 342 TTATACCTAGCTCTCAATAGCCCGGGGAGATGACTACCCCGTTCATGTCGTGGCG 401
 QY 301 CGAAAACGGAAGCGACCGGAATGTATCCACACAGTAATCTGTGTTATCATACGC 360
 DB 402 CGAAAACGGAAGCGACCGGAATGTATCCACACAGTAATCTGTGTTATCATACGC 461
 QY 361 TGACGCCAGAAATCAGTCAGCGGTCCTGTTAAAGCAGACAGCGGTTTCTTGT 420
 DB 462 TGACGCCAGAAATCAGTCAGCGGTCCTGTTAAAGCAGACAGCGGTTTCTTGT 521
 QY 421 TCCAGCTCTTTTGTGCTGAACATCGGGTAATCTGCTCTTAAACCAAGTAAATCGT 480
 DB 522 TCCAGCTCTTTTGTGCTGAACATCGGGTAATCTGCTCTTAAACCAAGTAAATCGT 581
 QY 481 TTTTITAGGTCCTGACACACGCTGGCAGTAGCGTATTGTGGCACAAAATAGAC 540

DB 582 TTTTITAGGTCCTGACACACGCTGCGACAGTAGCGTATTGTGGCACAAAATAGAC 641
 QY 541 ACACCGGAGTTTCATCATGACCTTAGAATGGTGGTTTGGCTTACCTGCTGACATCGATCAT 600
 DB 642 ACACCGGAGTTTCATCATGACCTTAGAATGGTGG-TTGGCTTACCTGCTGACATCGATCA- 699
 QY 601 TTTAACGGTGTGGCAGGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCG 650
 DB 700 TTTAACGGTGTGGCAGGCTCTGGTGAATCAACACTATGACCACTCGCTCAACACCG 759
 QY 661 TTAATC 665
 DB 760 GTATC 764

RESULT 2

AQ859250/c

LOCUS

DEFINITION

AQ859250.1 GI:6209707

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1. (bases 1 to 473)

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

JOURNAL

COMMENT

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAAACAGCTATGACCATG

Class: BAC ends

High quality sequence start: 41

High quality sequence stop: 415.

Location/Qualifiers

1. .473

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/strain="Japonica"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone="nbe001114f"

/tissue_type="Leaf"

/lab_host="E. coli DH10B"

/clone_lib="CUGI Rice BAC Library (EcoRI)"

/note="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;

Rice is the most important food crop in the world. Half of

the world population, especially those inhabiting highly

populated areas of the humid tropics and subtropics, rely

on rice as their primary source of carbohydrate.

Monocotyledonous rice is a diploid plant (2n=24) with a

haploid genome equivalent of 431 Mbp (Arumuganathan and

Earle, 1991). The relatively small genome of rice, three

times larger than that of Arabidopsis, makes it suitable

for genomic studies. In order to facilitate positional

cloning, physical mapping and genome sequencing of rice,

we have constructed a BAC library from Oryza sativa,

Nipponbare variety using EcoRI as the cloning enzyme. The

library contains 55,296 clones with an average insert size

of 121 Kb providing approximately 15 haploid genome

equivalents. The deep coverage allows the isolation a

particular sequence with a probability of 99.9%. Three


```
/organism="Anopheles gambiae"
/mol_type="genomic DNA"
/strain="PEST"
/db_xref="taxon:7165"
/clone="18P02"
/clone_lib="Notredame1"
/notes="end : T7"

ORIGIN
Query Match      18.2%; Score 218.2; DB 29; Length 815;
Best Local Similarity 77.0%; Pred. No. 5e-55;
Matches 298; Conservative 4; Mismatches 82; Indels 3; Gaps 3;

QY 1 AGAAATAATGTGAGATCGACGCGCCATCGAATGTGCGAGTATATAGCGTTTACGCCAC 60
Db 423 AAAAAACAGCGTTGAAATCGCACCGCCATGAGTGGCGGCAATATAGCGTTCCGCCAG 482
QY 61 GGACCGGGGTGAACCTCTGCTGCCAGATGCGGCGAGATCATCAACATAATCAATTAAG 120
Db 483 GGGCTTGGCTGACCTCTGCTGCCAGACGCGGCGAGATCGTCGACGTAATCGCTGAG 542
QY 121 CGATTACATGCCGAGATCGGATCGGTTCACAGCGACCGGAACGTCCTGCGCCGCG- 179
Db 543 TTATCACGTCGGCGGATGGGTATCCGATACAGCGCGCCCTGAAACCCCTGCGCCGCGS 602
QY 180 ATGTCGATGATTAGACATCAACCCCAATGGAACAGGTATAGCGGAGTTCGGGATA 239
Db 603 GTGATCGAATACGACATCGAGCCGAGATGGAACAGGTATAGCGGAGTTCGGGATA 662
QY 240 TTTTACGTAGCTCTCAATACGCGCGCGGAGATGACTACACCGGTGATGCTGTGTC 299
Db 663 TTTTATGTAGCTCTCAATACGTCGCGGCGAGACCAATGACCGGTGATTTTCTTATC 722
QY 300 GCGAAACGAGACAAAGCGGACGATGTCATCCACACGAGTAACTGCTTCTCATCAG 359
Db 723 GTG-AAAGCACAAGCGTACTGGAATATCGCGGAGCGCCGTAATCTTCTCTCAG 781
QY 360 CTGACCGCGAATATGATCAGCGGTC 386
Db 782 CTG-CGCTAGAAATCGTCAGCGGCC 807

RESULT 5
AF114214 954 bp DNA linear GSS 29-AUG-2000
LOCUS
DEFINITION
AF114214 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 415-T3, genomic survey sequence.
ACCESSION
AF114214
VERSION
AF114214.1 GI:5281261
KEYWORDS
GSS.
SOURCE
Salmonella typhimurium
ORGANISM
Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 (bases 1 to 954)
AUTHORS
Wong R.M.Y. and McClelland M.
TITLE
End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,
Li-Cor
JOURNAL
Unpublished (1999)
COMMENT
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifesci.sdsu.edu
Class: shotgun.
Location/Qualifiers
1..954
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="LT2"
/db_xref="taxon:602"
/clone="415-T3"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"

/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_1079"
/clone_lib="pacs1-60"
/notes="clinical isolate 1-60 Whole genomic shotgun
library."

ORIGIN
Query Match      7.1%; Score 85.4; DB 28; Length 1095;
Best Local Similarity 54.1%; Pred. No. 3.2e-14;
Matches 217; Conservative 0; Mismatches 181; Indels 3; Gaps 2;

QY 774 TGTGAAGTGGGAGCGCGGCTTACTTGTGTTGGTGGGAATCCAGAGTGGC--GCGC 831
Db 41 TGATCAAGTGGTTCGGGTGCGCTTACTTGTGTACTTGGCGGCGGCGGCGGCGC 100
QY 832 CGCTGGTGCATTTACCTTAATCGTGGGCTTACTCAATCG-CGTCGACATTTCTCC 890
```

```
/note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"

ORIGIN
Query Match      10.0%; Score 120.6; DB 28; Length 954;
Best Local Similarity 77.0%; Pred. No. 4.7e-25;
Matches 147; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1 AGAAATAATGTGAGATCGACCGCCATCGAATGTGCGAGTATATAGCGTTTACGCCAC 60
Db 763 AGGAACAGGGTGGCTATCGCGCCCGCCATGAGTGGCGGAGATATACGCTTACGCCAC 822
QY 61 GGACCGGGGTGAACCTCTGCTGCCAGATGCGGCGAGATCATCAACATAATCAATTAAG 120
Db 823 GCGCCCGGCTCAATTTCTGCTGCCAGACGCGGCTAAGTCTCAACATAATCGTTAAAA 882
QY 121 CGATTACATGCCGAGATGCGGATCGGCTAAACAGGCGACCGGAACGTCCTGCGCCGCA 180
Db 883 TGATCGACATGACCGCGGATGCGGTCGGGTAACATCGCGCGGAGCGCCCTTGACCACGG 942
QY 181 TGTGCGATGAT 191
Db 943 TGTGCGATGAT 953

BZ548406 1095 bp DNA linear GSS 17-DEC-2002
pacs1-60_1079.s1 pacs1-60 Pseudomonas aeruginosa genomic clone
pacs1-60_1079, genomic survey sequence.
BZ548406
VERSION
BZ548406.1 GI:27151987
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1095)
AUTHORS
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol. (2002) in press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1..1095
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_1079"
/clone_lib="pacs1-60"
/notes="clinical isolate 1-60 Whole genomic shotgun
library."

ORIGIN
Query Match      7.1%; Score 85.4; DB 28; Length 1095;
Best Local Similarity 54.1%; Pred. No. 3.2e-14;
Matches 217; Conservative 0; Mismatches 181; Indels 3; Gaps 2;

QY 774 TGTGAAGTGGGAGCGCGGCTTACTTGTGTTGGTGGGAATCCAGAGTGGC--GCGC 831
Db 41 TGATCAAGTGGTTCGGGTGCGCTTACTTGTGTACTTGGCGGCGGCGGCGGCGC 100
QY 832 CGCTGGTGCATTTACCTTAATCGTGGGCTTACTCAATCG-CGTCGACATTTCTCC 890
```

Db 101 CGCCACAGCCCTTGAGCACCAGTGGGACCGCTCTGGGGCGACCGTTGACCTGGTGC 160
QY 891 AGCGCCAGCTTTTGTGAATCTACCAATCCCAAGATTTGTGTTCTGGCGCGCTAT 950
Db 161 TGGTGGTTTCTGTGTGCAACGCGCAGCAATCCCAAGCGGTGATCTTCATGCTCGCGTGC 220
QY 951 TTCCGCAATTCATCATGCGCGCAACGCGCACTGATGAGTATATCGTGTCTGGCGTCA 1010
Db 221 TGGCGAGTTTCATCGACCGCGCAGCGCGCTGCTGGCGCAATACCTGATCATGGCGGCA 280
QY 1011 CCACATTTGTTGATATATTTGATGATCGGTACGCGCACCTTGTCTCAAGGATTTG 1070
Db 281 CCATGATCGTCTGACCGTATCGTATGCGCGGTACACCGCGCTGCTGCGCGCTGC 340
QY 1071 CTCTATGGATTAAAGGACCAAGACGAGTGAAGCGCTCGAATAAGATTTTGGCTCGTTGT 1130
Db 341 TAGCGTACTGCTGCTGCGCGCGCAGCAGAGCTGGTGAACCGTACCTTCCGACCGCTGT 400
QY 1131 TTATGCTGTGGAGCGCTGTGTAGCATCGCGAGCATGCG 1171
Db 401 TCGTGGTCCGCGGGCTGCTGGCGAGCGGTACGCGAGCG 441

RESULT 7
BZ566001/c
LOCUS
DEFINITION
pac82-164_5858.xl pac82-164 Pseudomonas aeruginosa genomic clone
pac82-164_5858, genomic survey sequence.

ACCESSION
BZ566001
VERSION
BZ566001.1 GI:27194705
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
1 (bases 1 to 1199)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
Whole-Genome-Sequence Variation among Multiple Isolates of
Pseudomonas aeruginosa Library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..1199
/organism="Pseudomonas aeruginosa"
/mol_type="Genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac82-164_5858"
/clone_lib="pac82-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN
Query Match 4.7%; Score 55.8; DB 28; Length 1199;
Best Local Similarity 50.1%; Pred. No. 4e-05;
Matches 195; Conservative 0; Mismatches 187; Indels 7; Gaps 2;

QY 584 GCTGGGCTTCAGACCGGATCGCGATTATTTGCTGCTGGTGGGCTGGGACGC 743
Db 649 GCTGGGCTTCAGATCGGGTGGCGCTCCCTGACAGATACCATTTCTCTGGCGGGTGGGTACG 590
QY 744 TATTTTCC--CGTCAGTGTGCGTTTGAAGTGGGCGGCGCGGCTTACTT 801
Db 589 TTAGTGGCGCCGCTCGGCACCTGGCTACAGCATGATCAGTGGTTCGCGGTGGCTACAT 530

QY 802 GATTTGGCTGGGAATCCAGCAGTGGC-----GGCGCGCTGGTGAATTGACCTTAATCG 856
Db 529 GGTGTACTGCTGTGGTGGCCAGTGGCAGGCGCCGCCACAGACCTTGAGCAGCGATGGCG 470
QY 857 CTGGCCTCTACTCAATCGCGTCGACATTTGTCAGCGCGCAGTTTGTGAATCTCACC 916
Db 469 ACCGCGCCTCTGGGTGCGACCGTTGACCTGTGCTGGTGGTCACTGGTCAGCGACACC 410
QY 917 RATCCCAAAAGTATTTGTTCTGCGCGCGCTATTTCCGGAATTCATCATGCCGCAACAG 976
Db 409 AATCAACAAGACGGTGTCTGCAATGCTCGCGGTGGTCCCCAGATTTCAGCGACCGCATCAT 350
QY 977 CGCAACTGATGTCAGTATATATGCTGCTGGCGGTACCACTATTGTGGTGCATATTTATTTG 1036
Db 349 CGCTGTCTGGCGCAAGACCTGATCATGCGCGGCCCATGATCGTGGTGGTGGTATCGTGC 290
QY 1037 ATGATCGGTTAGCGCACCGCTTGTCTCAAG 1065
Db 289 ATGTCGAGTACACCGGGCTGGCTGCGCG 261

RESULT 8
CB042356/c
LOCUS
DEFINITION
4007198 BARC-EMBRAPA 316BOV Bos indicus cDNA clone 316BOV_1P12
Unknown, mRNA sequence.

ACCESSION
CB042356
VERSION
CB042356.1 GI:27761601
KEYWORDS
EST.
SOURCE
Bos indicus (zebu)
ORGANISM
Bos indicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE
1 (bases 1 to 231)
Wood, D.L., Capuco, A.V., Brito, M.A.P., Martinez, M.L., Connor, E.E.,
da Mota, A.F., Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K.,
Machado, M.A. and Coutinho, L.L.
Construction and Characterization of cDNA Libraries Generated from
Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
indicus) Cattle
Unpublished (2002)
Contact: Adilson F. da Mota
Gene Evaluation and Mapping Laboratory
USDA, ARS, Animal and Natural Resources Institute
Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
Tel: 3015048456
Fax: 3015048414
Email: amota@npgl.embrapa.br
Single pass sequencing. Bases called and trimmed with phred
0.00925 using options -trim.alt'', -trim.fasta. Vector identified
by cross_match using options -minmatch 12 -minscore 18

PCR Primers
FORWARD: GTTTTCCAGTCACGAGTTG
BACKWARD: TGAGCGGATACAAATTCACAG
Plate: 1 row: P column: 12
Seq primer: GTTTTCCAGTCACGAGTTG
High quality sequence stop: 231.
Location/Qualifiers
1..231
/organism="Bos indicus"
/mol_type="mRNA"
/strain="Brazilian Dairy Gir"
/db_xref="taxon:9915"
/clone="316BOV_1P12"
/sex="female"
/tissue_type="mammary"
/cell_type="epithelium"
/dev_stage="involved"
/lab_host="K-12"
/clone_lib="BARC-EMBRAPA 316BOV"
/note="Organ: mammary; Vector: pUC19; Site 1: SmaI;
Site 2: SmaI; This mammary-derived cDNA library was

created as part of a collaborative project between the ARS Gene Evaluation and Mapping Laboratory and the EMBRAPA Dairy Cattle Research Center under the sponsorship of USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA extracted on 6/8/02. RT with Superscript II at 37 deg C, annealing temperature, PCR with 16mer AM5."

ORIGIN

Query Match 4.4%; Score 53; DB 14; Length 231;
Best Local Similarity 56.6%; Pred. No. 0.00012;
Matches 98; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 878 CGACATTTGTTCCAGCGCGAGTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTT 937
Db 201 CGCGTCTGTTCTGTTGATTGTCACCAATCTGACGACCGGACCGATCTGTTTC 142
QY 938 CTGCGCGGGTATTTCGGCAATTCATCTGCGCAACAGCGCGCAACTGATGAGTATATC 997
Db 141 ATGTTGCGCGTCTGCGCGAGTTCATCGATCGGCGCAACCGCTGTTTCGCAACTCGGC 82
QY 998 GTGCTCGCGGTCACCACTATTGTTGGTATATTTGATGATCGGTTATCC 1050
Db 81 ATCTGCGCGCGAGTGTGTTCTGCGACCTGATGTCATGCGGCTATGC 29

RESULT 9

AQ991154
LOCUS 667 bp DNA linear GSS 14-AUG-2000
DEFINITION Rf02035 Photorhabdus luminescens strain W14 M13 library
Photorhabdus luminescens genomic clone PLG02035, genomic survey sequence.
ACCESSION AQ991154
VERSION AQ991154.1 GI:9649748
KEYWORDS GSS.
SOURCE Photorhabdus luminescens
ORGANISM Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.
REFERENCE 1 (bases 1 to 667)
AUTHORS ffrrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T., Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium Photorhabdus luminescens W14: potential implications for virulence
JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE 20378633
PUBMED 10919786
COMMENT Contact: ffrrench-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bssr@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see ffrrench-Constant et al. 2000, Nucleic Acids Res.
Seq primer: M13 Forward
Class: shotgun.

FEATURES

Location/Qualifiers
source
1..667
/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/strain="W14"
/db_xref="taxon:29488"
/clone="PLG02035"
/dev_stage="primary phase variant"
/clone_lib="Photorhabdus luminescens strain W14 M13 library"
/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

ORIGIN

Query Match 3.7%; Score 44.8; DB 28; Length 667;
Best Local Similarity 55.6%; Pred. No. 0.069; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 69;
QY 25 CCATCGAATGTGCCAGTATATAGCGTTTACGCCACGCGCGGTGAACCTCTGCTGC 84
Db 505 CCCATTGAATGGCTTAACGATACAGTGAATAATGACGAGGAAGATTCTTTTCG 564
QY 85 CAGAATGCCCGCAGATCATCAATCAATTAAGCGATTAAATGATGCCGAGATGGGA 144
Db 565 ATAAATATTTCACAGTCATTATATAGTATCAATATTTTCTACATGCCCTTTTGTGA 624
QY 145 TCGGCTTAACAGCGCGACCGGAAAGTCCCTGCCCG 177
Db 625 TCTGCAAGCATCTCTGTGATCGCTGCCACG 657

RESULT 10

CNS014YP
LOCUS 922 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN12B16 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL104779
VERSION AL104779.1 GI:5616793
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 922)
AUTHORS Direct Submission
TITLE Drosophila melanogaster (fruit fly)
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

Location/Qualifiers
source
1..922
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN12B16"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"

ORIGIN

Query Match 3.4%; Score 41.4; DB 29; Length 922;
Best Local Similarity 24.3%; Pred. No. 0.91;
Matches 73; Conservative 95; Mismatches 133; Indels 0; Gaps 0;
QY 303 AAACGGACAAGCGCACCGGAATGTCATCCACACGAGTAACCTCTGTTTATCAGCGTG 362
Db 568 AMWYKWTCTAAHYWCTMTCATATBCTTYCCCAAYMAATCTACTYCKTYMYCMY 627
QY 363 AGCCAGAAATCAGTCAGCGGTCCATGTTAAAGCAGCAACCGTTTCTCTTGTTC 422
Db 628 ACNAHCWTATCTHWWTCYTTBGYDKAAMWAMYSINMAAGATYRTATCACYCBATY 687
QY 423 CCAGTCTTTTCTGCTGAACATCGGGTATATGCTCTTTAAACCGAGTAAATCGTTT 482
Db 686 YACAYCTAYBTWTATATCBHCKEAWACTKDYVYTTARAHAMAAMWACBCCATT 747
QY 483 TTTTAGCGTGCCTCACACAGCGTGCAGTAGCGTATTGTGGCACAAAATAGACAC 542


```

Db      748 WHYTWYCCYGCYCGRAAAAWBKNTRGASSARRCKGCKVBBSKSAVNYRDVAAMCSC 807
QY      543 ACCGGAGATTTCATGACCTTAGAATGGTGGTTTCCCTACCTGCTGCATCATGATTT 602
Db      808 SAMCCCCMHRCKCYCWNARFADAGADMYSCBKTADBKTRKWMGKSGTAMCKCCCCMYK 867
QY      603 T 603
Db      868 T 868

RESULT 11
EX462111/c
LOCUS   EX462111
DEFINITION Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CS0DG001YP07 5-PRIME, mRNA sequence.
ACCESSION EX462111
VERSION   EX462111.1
KEYWORDS EX462111.1 GI:31029391
SOURCE   EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 849)
Li W.B., Gruber C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segret@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9966.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DG001CH04QP1&cluster=9966.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DG001CH04QP1.
Location/Qualifiers
1. 849
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG001YP07"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo (AT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match 3.4%; Score 41; DB 13; Length 849;
Best Local Similarity 21.3%; Pred. No. 1.2;
Matches 53; Conservative 86; Mismatches 110; Indels 0; Gaps 0;

QY      661 TTATCCGGCCGGTGGCGTCTATTGCTGGGCTTCAGACCGGATCGGCGATTATATTTGTC 720
Db      500 TTGTGCTGCTTTTCGCCGTCTTGTCTTAGCTTTGGCGCTATTGCGCGCGCATTTTGC 441
QY      721 TGGTTGGCGTGGGCTTGGGACGCTATTTCGCCCTCAGTGTGCTTTGAAGTGTTCGA 780
Db      440 TTGCKTGCKKCKGCKKCKKCKGCKTGCKGCKKGGCTSTTTSKKSKKSKGSKGS 381
QY      781 GTGGCAGCGCGCGCTTACTTGATTGGCTGGGAAATCCACAGATGGCGCGCGCTGTGTC 840
Db      380 KKSSTASGSGTSGSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSKKSGNS 321
QY      841 AATTGACCTTAAATCGCTGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGT 900

```

```

Db      320 TGSNCNCKSKAAKTATGSGSKGKGTASTTSTKKKSKSKKGSSTKKKKSKSKSKKA 261
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      901 TTTTGTA 909
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      260 CKGSKKKA 252

RESULT 12
CN5016W
LOCUS   CN5016W
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN16011 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106730
VERSION    AL106730.1 GI:5623541
SOURCE     GSS
ORGANISM   Drosophila melanogaster (fruit fly)
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1039)
AUTHORS    Genoscope.
TITLE      Direct Submission
JOURNAL    Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT    Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelosAC11.

FEATURES             Location/Qualifiers
     1..1039
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /clone="BACN16011"
        /clone_lib="DrosBAC"
        /plasmid="pBelosAC11"
        /note="end : T7"

ORIGIN
Query Match      3.4%; Score 40.6; DB 29; Length 1039;
Best Local Similarity 30.6%; Pred. No. 1.7;
Matches 57; Conservative 54; Mismatches 75; Indels 0; Gaps 0;

QY      628 AATCAACACTATGACCACTCGCTCAACACGGTATCCGCCCGGTGCGTCTATTGCTG 687
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      829 WAAAAMAAMAAMWTHMARRTGMAWNTTGGGGGTTTGGKGGTGTGGMMTTCMVGSMTM 888
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      688 GCGTTTCAGACCGGACCTGGCGATTTCATTTGCTGGTTGGGCTGGGGTTGGAGCGCTATT 747
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      889 KGTTTBBGVGGTTTGTGTVKTKHDKTGBTGCCCTTBYBCKGSGGTTGKYCYCCCYT 948
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      748 TTCGGCTCAGTGAATTCGGTTTGAAGTGTGAAGTGGGACGCGCGCTTACTTGATTG 807
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      949 BYCGGGKTTTTCCTGTTGGGKTKGGGTKKGGGGKGTTCNKKKCKCVHHTTTKKKKKKKKK 1008
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY      808 GCTGGG 813
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      1009 KTKGG 1014
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 13
BH209707/c
LOCUS    BH209707
DEFINITION Sml-41P23.TF Sml Schistosoma mansoni genomic clone Sml-41P23,
genomic survey sequence.

```

ACCESSION BH209707
VERSION BH209707.1 GI:16388592
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
REFERENCE Strigeidida; Schistosomatoidea; Trematoda; Digenea; Schistosoma.
AUTHORS Shetty, J., Simpson, A., Malek, J., Koo, H., LoVerde, P.T. and El-Sayed, N.M.
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sml BAC library for gene discovery and map construction
JOURNAL Unpublished (2001)
COMMENT Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Schistosoma mansoni (Puerto Rico) Sml BAC library. For clone availability, please contact Dr. Najib El-Sayed at TIGR (nelsayed@tigr.org) or Dr. Phillip LoVerde at State University of New York, Buffalo, New York, USA (loverde@buffalo.edu)
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1. .475
/organism="Schistosoma mansoni"
/mol_type="genomic DNA"
/strain="Puerto Rico"
/db_xref="taxon:6183"
/clone="Sml-41P23"
/clone_lib="Sml"
/notes="Vector: pBeloeBAC11; Site 1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBeloeBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 X the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."

Query Match 3.3%; Score 39.8; DB 28; Length 475;
Best Local Similarity 46.8%; Pred. No. 1.9;
Matches 125; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 933 TGTTCTGGGGGGCTATTTCGCAATTCATCATGCGCGCAACAGCGCAACTGATGCAGT 992
Db 352 TGATCTGCTGTTGCTGATGCTGCTGCTGATGATGCTGCTGCTGATGCTGCTGTC 293

QY 993 ATATCGTCTGCGCGTCACACATATTGCGTGCATATTATTTGATGATCGGTTACGCCA 1052
Db 292 TGCTGCTGCTGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGA 233

QY 1053 CCCTTGCTCAACGATGCTCTATGATGATTAAGGACCAAGCAGATGAAGCGCTGAATA 1112
Db 232 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATA 173

QY 1113 AGATTTTGGCTGCTGTTTATGCTGCTGGAGCGCTGTAGCATCGCGAGGCATCGGT 1172
Db 172 AGATTTTGGACATGATGCTGTTTGTGACCATGACGATGATGATGATGATGATGATGA 113

QY 1173 GAAAAAATGTCGGATCGCGCTAAA 1199
Db 112 GGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATA 86

RESULT 14
AY401573
LOCUS Mus musculus GPR7 gene, VIRTUAL TRANSCRIPT, partial sequence, GSS.
DEFINITION Genomic survey sequence.
ACCESSION AY401573
VERSION AY401573.1 GI:39757562
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 990)
AUTHORS Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
Location/Qualifiers
1. 990
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>990
/gene="GPR7"
/locus_tag="HCM0935"
ORIGIN
Query Match 3.3%; Score 39.8; DB 29; Length 990;
Best Local Similarity 52.1%; Pred. No. 2.9;
Matches 112; Conservative 0; Mismatches 102; Indels 1; Gaps 1;

QY 580 CTACTCTGTCGATCGATCATTTTAAAGCTGTCGCCAGGCTCTGGTGCATCAACACTAT 639
Db 337 CTCATTGTAGCGCTGACACAGTACACATCTTCTAGCTCTACTTC-CTCGCGTCAT 395

QY 640 GACCACTTCGCTCAACACAGGTTATCCGCGCGGTGGCGTCTATTGCTGGGCTTCAGACCG 699
Db 396 GAGCGCCGACGACATACCTGTTCTTGGCCACAGAGTTCGCGCGGTGTCGCGGCG 455

QY 700 GACTGCGGATTCATATTGCTGCTGCTGCGTGGGTTGGACGCTATTTTCCGCTCAGT 759
Db 456 CACTTACCGTTCAGCGCGTGTCTGTCAGTCTGGCGGTGTGGCGCTGCTGCTGCTGCT 515

QY 760 GATTCGCTTTGAAGTGTGAAAGTCGGCAGCGCGG 794
Db 516 GCTGCCCTTTGCGGTATTTCGCTCGGCTGGACGAGG 550

RESULT 15
BX416727
LOCUS Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone EST. CS0DA011114 5-PRIME, mRNA sequence.
DEFINITION
ACCESSION BX416727
VERSION BX416727.1 GI:30765629
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Search completed: March 28, 2004, 18:21:15
Job time : 3830.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:04:39 ; Search time 2572.83 Seconds
(without alignments)
10360.543 Million cell updates/sec

Title: US-09-847-392-1_COPY_557_1171
Perfect score: 615
Sequence: 1 atgaccttagaatggtggtt.....tagcatggcgagcatgcg 615

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_ph.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	615	100.0	1200	6	AR172781	AR172781 Sequence
2	615	100.0	1200	6	E41573	E41573 DNA encodin
3	615	100.0	1231	6	E48927	E48927 Novel gene
4	615	100.0	1231	6	AX030175	AX030175 Sequence
5	604	98.2	91414	1	ECOW85	M87049 E. coli gen
6	595	96.7	11509	1	AE000458	AE000458 Escherich
7	577.6	93.9	10253	1	AE015396	AE015396 Shigella
8	577.6	93.9	290976	1	AE016990	AE016990 Shigella
9	571.2	92.9	318703	1	AP002567	AP002567 Escherich
10	566.4	92.1	301566	1	AE016769	AE016769 Escherich
11	553.8	90.0	10592	1	AE005613	AE005613 Escherich
12	467	75.9	212936	2	AC020970	AC020970 Mus muscu
13	449.8	73.1	256373	2	AC020870	AC020870 Mus muscu
14	414	67.3	417	6	AR352333	AR352333 Sequence
15	379.2	61.7	258050	1	AL627278	AL627278 Salmonell
16	379.2	61.7	299991	1	AE016845	AE016845 Salmonell
17	375	61.0	645	6	AR384166	AR384166 Sequence
18	374.4	60.9	21692	1	AE006884	AE006884 Salmonell
19	374.4	60.9	96086	1	STYSTM01	AF233324 Salmonell
20	358.8	58.3	265383	2	AC020874	AC020874 Mus muscu
21	274.8	44.7	10957	1	AE013640	AE013640 Versinia
22	274.8	44.7	199050	1	AJ414159	AJ414159 Versinia
23	241.4	39.3	268294	2	AC020885	AC020885 Mus muscu
24	207.4	33.7	263	6	AR352166	AR352166 Sequence
25	182.8	29.7	62274	2	AC020833	AC020833 Mus muscu
26	170.4	27.7	206758	2	AC016129	AC016129 Drosophil
27	162.6	26.4	299938	1	AE016800	AE016800 Vibrio vu
28	161	26.2	245560	1	AP005330	AP005330 Vibrio vu
29	151.4	24.6	288108	1	AP005083	AP005083 Vibrio pa
30	149.8	24.4	11305	1	AE004104	AE004104 Vibrio ch
31	123.6	20.1	3131	1	AB105408	AB105408 Photobact
32	112.4	18.3	10475	1	AE004937	AE004937 Pseudomon
33	109.4	17.8	110000	2	AC074221	Continuation (3 of
34	104	16.9	313518	1	AE016856	AE016856 Pseudomon
35	86.4	14.0	303226	1	AE016774	AE016774 Pseudomon
36	81.6	13.3	25260	1	EX640436	EX640436 Bordetell
37	80	13.0	123385	1	EX640452	EX640452 Bordetell
38	80	13.0	346359	1	EX640411	EX640411 Bordetell
39	75.4	12.3	663	6	AR318227	AR318227 Sequence
40	74.8	12.2	345783	1	AP003001	AP003001 Mesothizo
41	72.8	11.8	63847	2	AC101459	AC101459 Mus muscu
42	72.2	11.7	346897	1	AP002995	AP002995 Mesothizo
43	70.6	11.5	13301	1	AE005614	AE005614 Escherich
44	70	11.4	349737	1	EX572597	EX572597 Rhodopseu
45	61.8	10.0	11501	1	AE009276	AE009276 Agrobacte

ALIGNMENTS

RESULT 1
AR172781
LOCUS AR172781 1200 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 1 from patent US 6303348.
ACCESSION AR172781
VERSION AR172781.1 GI:17912272
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1. (bases 1 to 1200)
Livshits,V.Arkadiyevich., Zakataeva,N.Pavlovna.,
Aleoshin,V.Veniamovich., Balareova,A.Valentinovna. and
Tokmakova,I.Lyovna.
TITLE
DNA coding for protein which confers on bacterium escherichia coli

Escherichia coli and process for producing L-amino acid					
JOURNAL	Patent: JP 2000116390-A 1 25-APR-2000;				
COMMENT	AJINOMOTO CO INC OS Escherichia coli PN JP 2000116390-A/1 PD 25-APR-2000 PF 12-OCT-1999 JP 1999289777 PR 13-OCT-1998 RU 98118425 PI VITARI ARUKAEVICH RIVISHITSU,NATARIYA PAVUROVUNA ZAKATAEVA, PI VLADIMIR VENIYAMINOVICH ARYOSHIN,ARA VALENTINOVUNA BERARYO, PI IRINA RIVOVUNA TOKUMAKOVA PC C12N15/09,C07K14/245,C12N1/21,C12P13/06,C12P13/08/(C12N1/21, PC C12B1/19), PC C12N15/00 CC FH Key Location/Qualifiers FT CDS (557)..(1171).				
FEATURES	Location/Qualifiers source 1..1200 /organism="Escherichia coli" /mol_type="genomic DNA" /db_xref="taxon:562"				
ORIGIN	Query Match 100.0%; Score 615; DB 6; Length 1200; Best Local Similarity 100.0%; Pred. No. 1.9e+185; Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	ATGACCTTAGAATGGTGGTTTGCTACTCGTCGTGACATCGATCATATTAAACGCTGTCGCCA	60		
Db	557	ATGACCTTAGAATGGTGGTTTGCTACTCGTCGTGACATCGATCATATTAAACGCTGTCGCCA	616		
QY	61	GGCTCTGTCGAATCAACACTATGACCACCTCGCTCAACACCGGTTATCCGGCCGGTGGC	120		
Db	617	GGCTCTGGTGCATCAACACTATGACCACCTCGCTCAACACCGGTTATCCGGCCGGTGGC	676		
QY	121	GTCATTGCTGGGCTTCAGACCGGACTGCGGATT CATATTGTGTGGTGTGGCGTGGGGTT	180		
Db	677	GTCATTGCTGGGCTTCAGACCGGACTGCGGATT CATATTGTGTGGTGTGGCGTGGGGTT	736		
QY	191	GGGACGCTATTTCCCGCTCAGTGATTGGTTGGTTGAAGTGTTCAGTGGGACGGCGCGCT	240		
Db	737	GGGACGCTATTTCCCGCTCAGTGA TTGGCTTGAAGTGTTCAGTGGGACGGCGCGCT	796		
QY	241	TACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC AATTGACCTTTAAATCG	300		
Db	797	TACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGC AATTGACCTTTAAATCG	856		
QY	301	CTGGCCTCTACTCAATCGGTGCACATTTGTTTCAGCGCGCAGTTTTTGTAATCTCAC	360		
Db	857	CTGGCCTCTACTCAATCGGTGCACATTTGTTTCAGCGCGCAGTTTTTGTAATCTCAC	916		
QY	361	AATCCCCAAAAGTATGTGTTTCTGGCGCGCGCTATTTC CGCAAATTCATCATGCCGCAACAG	420		
Db	917	AATCCCCAAAAGTATGTGTTTCTGGCGCGCGCTATTTC CGCAAATTCATCATGCCGCAACAG	976		
QY	421	CCGCAACTGATGCAATATCGTGTCTGGCGGTCAACA CTATGTGTGTGATATATTGTG	480		
Db	977	CCGCAACTGATGCAATATCGTGTCTGGCGGTCAACA CTATGTGTGTGATATATTGTG	1036		
QY	481	ATGATCCGTTACGCCACCTTGCTCAACGGATTGCTCT ATGGAATTAAGGACCAAAACAG	540		
Db	1037	ATGATCCGTTACGCCACCTTGCTCAACGGATTGCTCT ATGGAATTAAGGACCAAAACAG	1096		
QY	541	ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTAT TATGCTGGTGGGAGCGCTGTTAGCA	600		
Db	1097	ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTAT TATGCTGGTGGGAGCGCTGTTAGCA	1156		
QY	601	TCGGCGAGGATCGG	615		
Db	1157	TCGGCGAGGATCGG	1171		

QY 241 TACTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTCGCAATTGACCTTAAATCG 300
DB |||||
QY 797 TACTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTCGCAATTGACCTTAAATCG 856
DB |||||
QY 301 CTGGCCTCTACTCAATCGCTGCACATTTGTTCCAGCGCGCAGTTTTCGTAATCTCACC 360
DB |||||
QY 857 CTGGCCTCTACTCAATCGCTGCACATTTGTTCCAGCGCGCAGTTTTCGTAATCTCACC 916
DB |||||
QY 361 AATCCAAAAGTATGTGTTCTGGCGCGCTATTTCCGCAATTCATCATGCGCAACAG 420
DB |||||
QY 917 AATCCAAAAGTATGTGTTCTGGCGCGCTATTTCCGCAATTCATCATGCGCAACAG 976
DB |||||
QY 421 CCGCAACTGATGTCAGTATATCGTCTCGCGCTGCACCACTATTCGTCGATATTATG 480
DB |||||
QY 977 CCGCAACTGATGTCAGTATATCGTCTCGCGCTGCACCACTATTCGTCGATATTATG 1036
DB |||||
QY 481 ATGATCGGTTACGCCACCCCTTGTCTCAACGATGCTCTATGGAATTAAGACCAAGCAG 540
DB |||||
QY 1037 ATGATCGGTTACGCCACCCCTTGTCTCAACGATGCTCTATGGAATTAAGACCAAGCAG 1096
DB |||||
QY 541 ATGAAGCGCTGATAAGATTTTCGGCTCGTCTGTTATGCTGGTGGCGCTGTTAGCA 600
DB |||||
QY 1097 ATGAAGCGCTGATAAGATTTTCGGCTCGTCTGTTATGCTGGTGGCGCTGTTAGCA 1156
DB |||||
QY 601 TCGCGAGGCGATGCG 615
DB |||||
QY 1157 TCGCGAGGCGATGCG 1171
DB |||||
RESULT 5
LOCUS ECOW85/c 91414 bp DNA linear BCT 29-MAY-1995
DEFINITION E. coli genomic sequence of the region from 84.5 to 86.5 minutes.
ACCESSION M87049
VERSION M87049.1 GI:836656
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 91414)
AUTHORS Daniels, D.L., Plunkett, G. III, Burland, V. and Blattner, F.R.
TITLE Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes
JOURNAL Science 257 (5071), 771-778 (1992)
MEDLINE 92358234
PUBMED 1379743
REFERENCE 2 (bases 1 to 91414)
AUTHORS Plunkett, G. III, Burland, V., Daniels, D.L. and Blattner, F.R.
TITLE Analysis of the Escherichia coli genome. III. DNA sequence of the region from 87.2 to 89.2 minutes
JOURNAL Nucleic Acids Res. 21 (15), 3391-3398 (1993)
MEDLINE 93347969
PUBMED 8346018
REFERENCE 3 (bases 1 to 91414)
AUTHORS Blattner, F.R., Burland, V., Plunkett, G. III, Sofia, H.J. and Daniels, D.L.
TITLE Analysis of the Escherichia coli genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes
JOURNAL Nucleic Acids Res. 21 (23), 5408-5417 (1993)
MEDLINE 94089392
PUBMED 8265357
REFERENCE 4 (bases 1 to 91414)
AUTHORS Rudd, K.E., Sofia, H.J., Koonin, E.V., Plunkett, G. III, Lazar, S. and Rouviere, P.E.
TITLE A new family of peptidyl-prolyl isomerases
JOURNAL Trends Biochem. Sci. 20 (1), 12-14 (1995)
MEDLINE 95184296
PUBMED 7878731
REFERENCE 5 (bases 1 to 91414)
AUTHORS Daniels, D.L.
TITLE Direct Submission

JOURNAL Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu; Phone: 608-262-2534; Fax: 608-263-7459
COMMENT On May 29, 1995 this sequence version replaced gi:148169. This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. The entire sequence was independently determined from E. coli M1665. Overlaps and conflicts with other sequence determinations are annotated. Reference [1] describes the original sequence determination of a 91408 bp sequence. References [2], [3], and [4] describe subsequent corrections and/or updates to that sequence.
FEATURES
source 1..91414
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/map="bp 1 at 3974 kb; 84.7 minutes"
/clone_lib="lambda library of Daniels and Blattner"
/notes="This sequence comprises the following lambda clones: DD850 (EC15-116), DD851 (EC13RM4), DD854 (EC15-177), DD856 (EC19RM8.1), DD858 (EC17RM9), DD861 (EC17-221), DD864 (EC27-149), DD865 (EC21-76), DD867 (EC27-860). M13mp19 or Janus vectors were used for subcloning."
misc_feature 1..5863
/note="clone EC15-116"
misc_feature complement(1..4899)
/note="corresponds to M37337 (1..4900)"
misc_feature 1..339
/note="corresponds to V00349 (164..500)"
misc_feature 1..200
/note="corresponds to K02846 (1..200)"
gene 10..86
/gene="aspt"
trna 10..86
/gene="aspt"
/note="CGSC No. 989"
gene 95..170
/gene="trpT"
trna 95..170
/gene="trpT"
/note="CGSC No. 66"
terminator 177..205
/note="terminator-like sequence"
misc_structure 253..396
/note="predicted bend of 81 degrees"
gene complement(266..862)
/gene="f198"
CDS complement(266..862)
/gene="f198"
/function="unknown"
/codon_start=1
/transl_table=1
/protein_id="AAA67566.1"
/db_xref="GI:148170"
/translation="MAHTRNENFSGASLASWECMLNQLGRLYQNDAHTGLQFEA
RIAQRSLVKQHERQLDLITTEAPKDEFSQLGLGYFTLYALTSAPSKLGLNLYL
RLEWGFDFQOHEAGLIGADEVPIITTSAAELAQOQIAMLNGCTWLPVSWARKKGLHT
VVDSTTLRPLAYIWLQNSDKNALIRDLKINVLDEVY"
misc_structure 586..724
/note="predicted bend of 77 degrees"
gene complement(703..1104)
/gene="f133"
CDS complement(703..1104)
/gene="f133"
/note="CGSC No. 18010; possible psr gene"
/codon_start=1
/transl_table=1
/protein_id="AAA67567.1"

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS
1 (bases 1 to 11509)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.

TITLE
JOURNAL
MEDLINE
The complete genome sequence of *Escherichia coli* K-12
Science 277 (5331), 1453-1474 (1997)
97426617

REFERENCE
AUTHORS
2 (bases 1 to 11509)
Blattner, F.R.

TITLE
JOURNAL
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE
AUTHORS
3 (bases 1 to 11509)
Blattner, F.R.

TITLE
JOURNAL
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
On Sep 9, 1997 this sequence version replaced gi:1790254.
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu/>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu/>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as Gene names.

FEATURES
source
1. .11509
/organism="Escherichia coli K12"
/mol_type="Genomic DNA"
/strain="K12"
/sub_strains="MG1655"
/db_xref="taxon:83333"
81. .950
/gene="pIdA"
/note="synonym: b3821"
81. .950
/EC_number="3.1.1.32"
/function="enzyme; Central intermediary metabolism: Phosphorus compounds"

gene
81. .950
/note="pIdA"

CDS
81. .950
/function="enzyme; Central intermediary metabolism: Phosphorus compounds"

/note="o289; 99 pct identical to PA1_ECOLI SW: P00631; CGSCNo. 384"
/codon_start=1
/transl_table=11
/product="outer membrane phospholipase A"
/protein_id="AAC76824.1"
/db_xref="GI:2367300"
/translation="MRTLQGLLPLVFMPLMAVYAQATVKEVHDAPVAGSIIANMLQ EHDNFTLYPDYTNLYITQTSIDLNKEATASDYAENARKDEVKFSLAFLPWRGIL NPSVIGASVTQKSNWOLSNSESPRETPNPQLPGFATPYAGWTLRVEVGY NHDNSGRSDPTSSNNPLYLRLAENGNMLVEVKPMVVGNTDNDPDKIKGYCYOLK IGYHLGDAVLSAKQYNNWNTGYGAGLGLSYPTIKHVRVLYTQVYSGESLIDYNFQ TRVGVGWMLNDLF"

promoter
965. .993
/note="factor Sigma70; predicted +1 start at 4003392"
1019. .1047
/note="factor Sigma70; predicted +1 start at 4003446"
1077. .2909
/gene="recQ"
/note="synonym: b3822"
1077. .2909
/gene="recQ"
/EC_number="3.6.1.-"
/function="enzyme; DNA - replication, repair, restriction/modification"
/note="o610; 99 pct identical to 607 amino acids of RECQ_ECOLI SW: P15043 but has 3 additional N-ter residues"

CDS
/codon_start=1
/transl_table=11
/product="ATP-dependent DNA helicase"
/protein_id="AAC76825.1"
/db_xref="GI:2367301"
/translation="MNVAAQEVNLESGAKVQLQETFGYQQRPGQEIIIDTVLSGRD CLVVNPTGGKSCYQIPALLNGLTVVSPILSMKQDQDQQLQANGAAACLNSTQT REQLEVMTQCRTPQLRIYIAPELRMLDNFLEHLAWNPVLAVDAECISQWGHF RPEYALQQRPTQLIPFALMTATADDTTQDQIVRLGLNDPLIQISSFPDRPIRYM LADVEKLDQIMRYVQQRGKGIYCNCRKAKVEDTAARLQSGISAAAYHAGLENNV RMQEFKDDQLQIVVAVFGWINKPNRVVHFDPNIESYQYGTGAGRDGL PABAMLYPDADMAWLRCLLEKQQLQDIEHKLNAAGAFACATCERVLVNLVYFG EGRPGQNCIDLPKQDQDQDQALSTIGTVNQRGQYVVEYVVGANNQIR DYGHDKLVQYMGKDSHEHVSIVIRQLHGLVTVNQIAHSAHQIETARFVLAESS LQAEPRVIALKPKAMQSGNYDRKLFAKRLKRSIADESNVPPYVVFNDATLIE MLEQMPTIASEMLSVNGVGRKLERFGKPFMALIRAFVDDDDDEE"

repeat_region
2104. .2140
/note="REP (repetitive extragenic palindromic) element; contains 1 REP sequence"
3162. .3189
/note="factor Sigma70; predicted +1 start at 4005588"
3224. .3592
/gene="yigU"
/note="synonym: b3823"
3224. .3592
/gene="yigU"
/function="orf; Unknown"
/note="G12; sequence change shortens and changes N-terminus relative to earlier version (YIGJ_ECOLI SW: P27846)"

promoter
3162. .3189
/note="factor Sigma70; predicted +1 start at 4005588"
3224. .3592
/gene="yigU"
/note="synonym: b3823"
3224. .3592
/gene="yigU"
/function="orf; Unknown"
/note="G12; sequence change shortens and changes N-terminus relative to earlier version (YIGJ_ECOLI SW: P27846)"

gene
3224. .3592
/gene="yigU"
/note="synonym: b3823"
3224. .3592
/gene="yigU"
/function="orf; Unknown"
/note="G12; sequence change shortens and changes N-terminus relative to earlier version (YIGJ_ECOLI SW: P27846)"

CDS
3224. .3592
/gene="yigU"
/note="synonym: b3823"
3224. .3592
/gene="yigU"
/function="orf; Unknown"
/note="G12; sequence change shortens and changes N-terminus relative to earlier version (YIGJ_ECOLI SW: P27846)"

repeat_region
3612. .3645
/note="REP (repetitive extragenic palindromic) element; contains 1 REP sequence"
3654. .4070
/gene="yigK"
/note="synonym: b3824"
3654. .4070
complement (3654. .4070)

/gene="yigK"
/function="orf; Unknown"
/note="f138; 100 pct identical to YIGK_ECOLI SW: P27847"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76827.1"
/db_xref="GI:1790258"
/translations="MIAFVFLKAGAAVLIWGIOWEAAAGAILDKSLASTQSRRLHP
ORAVFVNLNPKSIVFLAALFPQFIMPQPOLMQIVGVTVIIVVDIIIVMIGIATLAA
RIALWIKGKQKALNKIFGSLFVLGALLASARHA"
4200..4230
/notes="factor Sigma70; predicted +1 start at 4006629"
complement(4227..4254)
/notes="factor Sigma70; predicted +1 start at 4006612"
4384..5406
/gene="pidB"
/note="synonym: b3825"
4384..5406
/gene="pidB"
/EC_number="3.1.1.5"
/function="enzyme; Central intermediary metabolism;
Phosphorus Compounds"
/notes="0340; 99 pct identical amino acid sequence and
equal length to PLDB_ECOLI SW: P07000"
/codon_start=1
/transl_table=11
/product="lysophospholipase L(2)"
/protein_id="AAC76828.1"
/db_xref="GI:2367303"
/translations="WFOQKDMETRENAFAATMGPLTDFWRQDEAEFTGVDDIPVR
FVRPRAQHDVIVICPGRIEYKIVABLAYDLFLGDFVLIIDHROGGRSGLLADP
HLGPMFNDYVDDLAFAWQVQVQPKRYILAHSGGAIITFLQRIHPGVCDIAI
LTAPMGIVIRVDFVAFQKILNMAHPRFRDYGIAIGRWALFPAINLVTHSRQRY
RNLRFYADPTIRVGGTTPHWRVRESILAGVLAGAGDDATPTLLQAEERVVDNR
MDHRCFLATAAGHPVEGGRPLVKGAYHEILFEKDAMASVALHAIIVDFNRHNSPSG
NRSTEV"
5381..5408
/note="factor Sigma70; predicted +1 start at 4007807"
5426..5455
/notes="factor Sigma70; predicted +1 start at 4007854"
5438..5466
/notes="factor Sigma70; predicted +1 start at 4007865"
5494..5523
/note="factor Sigma70; predicted -1 start at 4007922"
5526..6041
Query Match
Best Local Similarity 98.9%; DB 1; Length 11509;
Matches 610; Conservative 0; Mismatches 5; Indels 2; Gaps 1;
QY 1 ATGACCTTAGATGGTGGTTCCTACCTGCTGACATCGATCATTTAACCTTCGCCA 60
DB 4273 ATGACCTTAGATGGTGGTTCCTACCTGCTGACATCGATCATTTAACCTTCGCCA 4214
QY 61 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTAT--CCGGCCGGTG 118
DB 4213 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGGTTATCGCGCGCGTG 4154
QY 119 GGCTCTATGCTGGGCTTCAGACCGGATTCATATGCTGCTGGTGGCGTGGG 178
DB 4153 GGCTCTATGCTGGGCTTCAGACCGGATTCATATGCTGCTGGTGGCGTGGG 4094
QY 179 TTGGGACGCTATTTCCCGCTCAGTGATTGCTGTTGAAGTGTGAAAGTGGCGAGCGCGG 238
DB 4093 TTGGGACGCTATTTCCCGCTCAGTGATTGCTGTTGAAGTGTGAAAGTGGCGAGCGCGG 4034
QY 239 CTTACTTTGATTTGGCTGGGAATCCAGCGTGGCGCGCGCTGGTGGCAATTGACCTTAAAT 298
DB 4033 CTTACTTTGATTTGGCTGGGAATCCAGCGTGGCGCGCGCTGGTGGCAATTGACCTTAAAT 3974
QY 299 CGCTGCTCTACTCAATCGCTGACATTTGTTCCAGCGCGAGTGTGTTGATCTCA 358

Db 3973 CGTGGCCTCTACTCAATCGCGCTGACATTTGTTCCAGCGCGAGTGTGTTGTAATCTCA 3914
QY 359 CCAATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATCGCGCAAC 418
Db 3913 CCAATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATCGCGCAAC 3854
QY 419 AGCCGCAACTGATGAGTATCATGCTGCTGGCGGCTACCACTATTGCTGCTGATATTATG 478
Db 3853 AGCCGCAACTGATGAGTATCATGCTGCTGGCGGCTACCACTATTGCTGCTGATATTATG 3794
QY 479 TGATGATCGTTTACGCCACCCCTTGTCTCAACGATTTGTCTATGATTAAGAGCAACAAGC 538
Db 3793 TGATGATCGTTTACGCCACCCCTTGTCTCAACGATTTGTCTATGATTAAGAGCAACAAGC 3734
QY 539 AGATGAGCGCGCTGAATAAGATTTTCCGCTGCTGTTTATGCTGCTGGCGAGCGCTGTAG 598
Db 3733 AGATGAGCGCGCTGAATAAGATTTTCCGCTGCTGTTTATGCTGCTGGCGAGCGCTGTAG 3674
QY 599 CATCGCGGAGGCGATCG 615
Db 3673 CATCGCGGAGGCGATCG 3657
RESULT 7
AE015396/c
LOCUS
DEFINITION Shigella flexneri 2a str. 301 section 359 of 412 of the complete genome.
ACCESSION AE015396
VERSION AE015396.1 GI:24054379
KEYWORDS Shigella flexneri 2a str. 301
SOURCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 10253)
AUTHORS Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
Liu,H., Yang,J., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
Wu,H.T., Dong,J., Sun,L.L., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,J.P.,
Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
Qiang,B.Q., Wen,Y.M., Hou,Y.D. and Yu,J.
TITLE Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157
JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)
PUBMED 12384590
REFERENCE 2 (bases 1 to 10253)
AUTHORS Jin,Q., Shen,Y., Wang,J.H., Liu,H., Yang,J., Yang,F., Zhang,X.B.,
Zhang,J.Y., Yang,G.W., Wu,H.T., Dong,J., Sun,L.L., Xue,Y.,
Zhao,A.L., Gao,Y.S., Zhu,J.P., Chen,S.X., Yao,Z.J., Wang,Y.,
Lu,W.C., Qiang,B.Q., Wen,Y.M. and Hou,Y.D.
Direct Submission
Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 YingXin Jie, XuanWu Qu, Beijing 100052, P.R. China
FEATURES
Location/Qualifiers
1..10253
/organism="Shigella flexneri 2a str. 301"
/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
complement(243..2167)
/gene="yigB"
/locus_tag="SF3893"
/psuedo
complement(243..2167)
/gene="yigB"
/locus_tag="SF3893"
note="similar to Escherichia coli K12 orf, conserved hypothetical protein gi: 1790248 (162 aa). BLAST with identity of 99% in 641 aa. This CDS has been disrupted by insertion. The sequence has been checked and is believed

```
to be correct."
/pseudo
/codon_start=1
/transl_table=11
472..1818
/locus_tag="SF3892"
472..1818
/locus_tag="SF3892"
/notes="Residues 1 to 448 of 448 are 98 pct identical to
residues 1 to 448 of a 448 aa protein from Shigella
flexneri ref: NP_085177.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN45328.1"
/db_xref="GI:24054380"
/translation="MFFDFMHGQALDVLVSYSRLNPLTSLGDLPELISRLAE
QRTVLRKRLPLEMVMGVIGVWLRKEPLHQIVNRDLIMLPGNRPVAPSATIOAR
QIGSEARVRFTKATQALWENATPHPHKGLTLLADGVFRTPTDPENDAAFPROTH
AGNALPYQVMVCMQELTSHLLTAAAFGTMKNSENLAELIETQNTLTLMDKGY
YSGLNWSLAGEHWHMILPRKGAQVEELRKLKGHDHLVKLTSPQARKWGLGN
EVARLLTVTRKVKVCHLLTMTDAMRFEGESMADLYSNREIELGYREIKQTWQLSR
TLRSKPELVEGELGVLLAYNLVRYQMIKVAEHLKGWPNQLSFSESCQWVRMLM
LQSGAPRIPELMSDLASGQLVKLPTRRGRAFRVVKRFPWKYPTAPKKSQSYA"
2522..3487
/genes="corA"
/locus_tag="SF3894"
2522..3487
/genes="corA"
/locus_tag="SF3894"
/notes="Residues 6 to 321 of 321 are 100 pct identical to
residues 1 to 316 of a 316 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_290445.1"
/codon_start=1
/transl_table=11
/product="MG2+ transport, system I"
/protein_id="AAN45329.1"
/db_xref="GI:24054381"
/translation="MGVPMVLSAFLQENNRRLTRLEVESQPLVNAVWIDLVEPDDBR
LRVQSGLSLATRPLEIEASRFEDDGLHHSFFFEADAEHAGNSTVAFTR
DGLFTRLELPAPFLYMRARSQSMVDGNAVYELLDLFTKLEQADENTYSOL
EQLSRIMEQHQDEYDEALSTAELEDIGKVRCLMDTORALNFLVRKARLPQGL
EQAREILLRDIETLIPNESLFOKNVPLQMAQMGFINIQNRILIKIESVSVVFLPPTL
VASSYGNPFEPPELXWSFGVGCALIFMLLAGLAPLYFKKNWL"
complement(3528..3950)
/genes="yigF"
/locus_tag="SF3895"
complement(3528..3950)
/genes="yigF"
/locus_tag="SF3895"
/notes="Residues 15 to 140 of 140 are 99 pct identical to
residues 1 to 126 of a 126 aa protein from Escherichia
coli K12 ref: NP_418261.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN45330.1"
/db_xref="GI:24054382"
/translation="MALLHQNSVGVPMKSKEYMNDLSLSEKWKYRNFVQHQFPQF
WGAIPYKFAKALKVRQLTTOHMFIAFCSWIYLFVLGLWKALIVLLGLHLSLEV
GALIGNIIAGIAVAAVYNTNKPVEKVGKLTWSL"
complement(3922..4362)
/genes="yigG"
/locus_tag="SF3896"
complement(3922..4362)
/genes="yigG"
/locus_tag="SF3896"
/notes="Residues 9 to 146 of 146 are 95 pct identical to
residues 1 to 138 of a 138 aa protein from Escherichia
coli K12 ref: NP_418262.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN45331.1"
/db_xref="GI:24054383"
/translation="WYESGHIAMLTGDGHKDVFKMLRFIPTSNKISRHRHYIELFIL
INFIAPFLIPNDGEAGFLVIVSTIALHVLVINNCORLDSGFIYIKIVVFTLAV
YIISITWTAEDEFACSGNSMELICYESTSMLMLAPTDSKQ"
complement(4397..5287)
/genes="radD"
/locus_tag="SF3897"
complement(4397..5287)
/genes="radD"
/locus_tag="SF3897"
/notes="Residues 1 to 296 of 296 are 99 pct identical to
residues 1 to 296 of a 296 aa protein RARD_ECOLI sp:
P27844"
/codon_start=1
/transl_table=11
/product="RARD protein"
/protein_id="AAN45332.1"
/db_xref="GI:24054384"
/translation="MDAKQTRQGVLLALAAVFIINGIAPAVEKLYVYVPADEILLTRVIL
NSFFPMVILMSICROWSYLKLITQTPQKILFAMVSAVLIGSNWLLFIWVANNHMLEA
SIGYINILNIVLIGMIFGERFRMQWLAIVLACGLVQLVQWTFGLSPLIAGLGAAS
FAPYGLVRKKIAVEAQGLIETMWLLEPVAAYIFAIADSTSHVGNQNSLNLILLA
AGIVTVFILLCTFAAATRLRLSTLGGFYQIYIGTLMFLAVTFYGEKEGADKXVTFIFI
WVALAIFVMDAIYTORTSK"
complement(5339..5824)
/genes="yigI"
/locus_tag="SF3898"
complement(5339..5824)
/genes="yigI"
/locus_tag="SF3898"
/notes="Residues 1 to 161 of 161 are 100 pct identical to
residues 1 to 161 of a 161 aa protein from Escherichia
coli O157:H7 ref: NP_312777.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN45333.1"
/db_xref="GI:24054385"
/translation="MKCKDDMSAVLTAEQALKVGENFVYHMPNRRALGMELERYEKE
PQALAFKQPMVGNWAQOSILHGGVIAALDVAAGLVCVGTSTFRHETISEDELURQL
SMQGTIDRLVDYLRPGRGERTAISSLLRAGNKVAAVAVELHNEEQLYIASATATYVY
C"
5971..6840
/genes="pldA"
/locus_tag="SF3899"
5971..6840
/genes="pldA"
/locus_tag="SF3899"
/notes="Residues 1 to 289 of 289 are 100 pct identical to
residues 1 to 289 of a 289 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_290453.1"
/codon_start=1
/transl_table=11
/product="outer membrane phospholipase A"
/protein_id="AAN45334.1"
/db_xref="GI:24054386"
/translation="WRILQGLLFPVFLPMNAVYAQEAIVKEVHDAPAVRGSIIANMLQ
EHDNFTIYDITNYLIYTQSDLNKEAIASDYMAENARKDEVKQSLAPFLWRGIL
GPNVLGYASDYTKSQWQLSNSESSPNFNPQFLGFATDYRFAGWTLRDVMGY
NHDNSGRSDPTSRSNRLYTRLMAENGNELVYKVPWYVGVNTDDNDPDTIKMGYQLK
IGYHLGDAVLGAKQYNWNTGYGAGLGLSYPITKHVRLYTQVYSGYGESLIDYFNQ
TRGVGVMLNDLF"
5988..8802
/genes="recQ"
/locus_tag="SF3900"
5988..8802
/genes="recQ"
/locus_tag="SF3900"
/notes="Residues 1 to 604 of 604 are 99 pct identical to
residues 8 to 611 of a 611 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_290454.1"
/codon_start=1
```

Query Match		93.9%;	Score 577.6;	DB 1;	Length 10253;
Best Local Similarity		98.1%;	Pred. No. 2.7e-173;		
Matches	Conservative	0;	Mismatches	9;	Indels
3;	Gaps	2;			
QY	1	ATGACCTTGAAGTGGTGGTTGCCCTACCTGCTGACATCGATCATTTTAAAGCTGTCGCCA	60		
Db	10168	ATGACCTTGAAGTGGTGGTTGCCCTACCTGCTGACATCGATCATTTTAAAGCTGTCGCCA	10109		
QY	61	GGCTCTGGTGCAATCAACTATGACACCTCGCTCAACCCAGGTTAT - CGGCCCCGTG	118		
Db	10108	GGCTCTGGTGCAATCAACTATGACACCTCGCTCAACCCAGGTTATCGGCGCGCGGTG	10049		
QY	119	GGGCTATTGCTGGGCTTCAGACCCGAGCTGGCGATTCTATTGTCTCGTTGGCGTGCGG	178		
Db	10048	GGGCTATTGCTGGGCTTCAGACCCGAGCTGGCGATTCTATTGTCTCGTTGGCGTGCGG	9989		
QY	179	TT - GGGACGCTATTTTCCCGCTCAGTGAATTGCGTTGGAAGTGTGAAGTGGCGAGCGCG	237		
Db	9988	TTGGGACGCTATTTTCCCGCTCAGTGAATTGCGTTGGAAGTGTGAAGTGGCGAGCGCG	9929		
QY	238	GCTTACTGATTGGCTGGGAATCAGACAGTGGCGGCGCTGGTGCAATTCAGCTTAAA	297		
Db	9928	GCTTACTGATTGGCTGGGAATCAGACAGTGGCGGCGCTGGTGCAATTCAGCTTAAA	9869		
QY	298	TCGCTGGCTCTACTCAATTCGGGTGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC	357		
Db	9868	TCGCTGGCTCTACTCAATTCGGGTGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC	9809		
QY	358	ACCAATCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGAA	417		
Db	9808	ACCAATCCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGAA	9749		
QY	418	CAGCGCAACTGATGCAGTATATCGTGTGGCGGTCAACACTATTGCGGTGATATTATT	477		
Db	9748	CAGCGCAACTGATGCAGTATATCGTGTGGCGGTCAACACTATTGCGGTGATATTATT	9689		
QY	478	GTGATGATCGGTATGCCACCCTTGCCTAAACCGATTGCTCTATGATTTAAAGACCAAAG	537		
Db	9688	GTGATGATCGGTATGCCACCCTTGCCTAAAGGATTGCTCTATGATTTAAAGACCAAAG	9629		
QY	538	CAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTA	597		
Db	9628	CAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGTGGGAGCGCTGTTA	9569		
QY	598	GCATCGCGGAGGATGCG	615		
Db	9568	GCATCGCGGAGGATGCG	9551		

RESULT 8	AE016990	290976 bp	DNA	linear	BCT 22-APR-2003
LOCUS	AE016990				
DEFINITION	<i>Shigella flexneri</i> 2a str. 2457T section 13 of 16 of the complete genome.				
ACCESSION	AE016990	AE014073			
VERSION	AE016990.1	GI:30042943			
KEYWORDS					
SOURCE	<i>Shigella flexneri</i> 2a str. 2457T				
ORGANISM	<i>Shigella flexneri</i> 2a str. 2457T				
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; <i>Shigella</i> .				
REFERENCE	1 (bases 1 to 290976)				
AUTHORS	Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunket, G. Jr., Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.				

TITLE	Complete Genome Sequence and Comparative Genomics of <i>Shigella flexneri</i> Serotype 2a Strain 2457T Infect. Immun. 71 (5), 2775-2786 (2003)
JOURNAL	2 (bases 1 to 290976)
PUBMED	Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Pournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Ferna, N.T., Payne, S.M., Rungen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
REFERENCE	Direct Submission
AUTHORS	Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
TITLE	Location/Qualifiers
JOURNAL	1. 290976
FEATURES	/organism="Shigella flexneri 2a str. 2457T"
source	/mol_type="genomic DNA"
	/strain="2457T"
	/serotype="2a"
gene	/db_xref="taxon:198215"
	complement(68..1405)
	/gene="yjdE"
CDS	/locus_tag="S3622"
	complement(68..1405)
	/gene="yjdE"
	/locus_tag="S3622"
	/function="putative transport; Not classified"
	/note="residues 1 to 445 of 445 are 92.35 pct identical to residues 1 to 445 from <i>Escherichia coli</i> K-12 : B4115"
	/codon_start=1
	/transl_table=11
	/product="putative amino acid/amine transport protein"
	/protein_id="AAP18666.1"
	/db_xref="GI:3004294"
	/translations="MSDSADAHKVLGIPTVLMVSGNMGSGVFLLPANLSTGGGAIY GMVLITIGALGSLVFPYKQKSLDIPSPGGYAVARRCFGLGQVTNLYVLACWIGNI AMVTVIGVGLYSFPFKLDFLTITCVVLMTFVLLNIVGPKMTTRVCATVLA PIVGLVFGHFMERGTWYAMNVSGLTGEGLIOSTLNTLMSFIGVESASVAAVGVK NPRNPVPIATIGVLTAAVCVYLSTIINGMTNPAAILRVASPPGDAARMALGIDTGA IVSFAAAGCIGLGGTLLAGTAAKADDGLFPFARVNAKGTGPFAGLIIIVGLIM TIFQLSGSNATKEFLGVSSSVITFLPYLTCAALLLGHGHGFKARPAYLAVIT IAFLYCIWAVVSGAKEVWMSFVTLMTVITVALNRYNLRHKKPYPLDAPISKD"
gene	complement(1542..2303)
	/gene="adiY"
	/locus_tag="S3623"
CDS	complement(1542..2303)
	/gene="adiY"
	/locus_tag="S3623"
	/note="residues 1 to 253 of 253 are 93.28 pct identical to residues 1 to 253 from <i>Escherichia coli</i> K-12 : B4115"
	/function="putative regulator; Not classified"

```

/codon_start=1
/transl_table=11
/product="putative ARAC-type regulatory protein"
/protein_id="AAP18667.1"
/db_xref="GI:30042945"
/translation="MRICSDQPCIVLLTERDKVMIRVNGKEPISLKNANHMALNCENNI
IDVSLNNTVAHSHIDIKVLRFLNKLSQIPYWRGSAITLTLPCLTDPVFRVA
KDNMPAETSEKERTRALLFTVLSRFLKFLSLMYMLRNCVSDSVYQIIESDIH
KDNVLSVACLCLSLKSLKKLKSNTYSQIITCRVYAVNLMMDGKNISQVSQ
SCGVNSTFISVFKDPFYGTPLHYVSQHRRTVA"
gene .
complement(2628..4895)
/gene="adia"
/locus tag="S3624"
complement(2628..4895)
CDS
/gene="adia"
/locus tag="S3624"
/function="enzyme; Degradation of small molecules: Amino
acids"
/note="residues 1 to 755 of 755 are 99.47 pct identical to
residues 2 to 756 of 756 from Escherichia coli K-12 :
B4117"
```

```

/codon_start=1
/transl_table=11
/function="biodegradative arginine decarboxylase"
/protein_id="AAP18668.1"
/db_xref="GI:30042946"
/translation="MKVLIVSEFELHQDTWGVNAVERLADALSQONVTVIKSTFDDG
FAILSNEAIDCLMFSYQWHEHONVROLIGKLHERQONVPVLLGDKREKALAAMD
RDLELVDFPALTEDTDFIAGRAVAATVROQLLPFLPSALMKYSDIHEYSWAAP
GQGGVFTKTAGRTHDYIYENLFRDIMGIERHTSLGSLDHTGAGSESEKYAARVP
GADRSWVGTSNRTIQACMDNDVVVDNRCHASIEQGLMLTGAKRFTVNPSPR
NRYGIIGPIYQEMOPELTKQKISSEPLTKKAGQKPSYCVVNTYDGVGCYNAAKEAQ
DLLEKTSRDLHDEAWYGVARENPFIYADHYAMRGEPGHNGFTVFATHSIHKLLNALS
QASYIHVRGEGALNPSRNQAMHATSPLYAI CASNDVAVSMMDGNSGLSTQV
IDSAVDFQOMARLYKEFTADGSEWFKPWKNKEVVDPTQKTYDFADAPT KLLTQVQ
CWMVHGBSWHGFKDIPDNMSMLDPIKVSILAPGNGEGELEEGVPAALVTAWLGRH
GIYPTTITPQINFPLSMGVTGKGTLVNTILCSFKHYDANTPLAQMPELVEQYPD
TYANMTHIDLOFNAWLKENNPGARMEAYSGLPVASEITPREAYNAIVDNNVELSI
ENLPGRIANSVIPYPPGIPMLLSGENTFGDKNSPQSVYLSLQSDWHDHFRFPHETEG
TEIIDGIYHMCVKA"
complement(5094..6002)
/gene="melR"
/locus_tag="S3625"
/locus_tag="(5094..6002)
/complement(5094..6002)
/gene="melR"
/locus_tag="S3625"
/function="regulator; Degradation of small molecules:
Carbon compounds"
/notes="residues 1 to 302 of 302 are 99.66 pct identical to
residues 1 to 302 of 302 from Escherichia coli K-12 :
B4118"
/codon_start=1
/transl_table=11
/product="regulator of melibiose operon"
/protein_id="AAP18669.1"
/db_xref="GI:30042947"
/translation="NNTDTFMCSSDBEKOTRSPLSLYSYORMEIEFRAPHIMPTSHWH
GOEVNAVPDGVDVYLINNEKINQGHITLFWACTPHQLDTDGTCCOSMALFNLPJLH
FLSWPDKDLINHVTGHWIKSLATQQLSPFEVRRWQELNSPEQIRQLAIDIGLM
LKRFSSWEPILNPKTSIRKNSVSRRAQFVYSQGLFIENIDQALTDINTVAEHVK
LNANYAMGIFQVWVMTKQYITAMRINHRVALLSDTDKSLDIALTAGFRSSSRFYS
TFQKYVYMSPPQYRKLQQRRQTFPG"
6060..6114
/notes="BOXC-20"
/gene="melA"
/locus_tag="S3626"
/locus_tag="(5094..6002)
/complement(5094..6002)
/gene="melA"
/locus_tag="S3626"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/notes="residues 1 to 451 of 451 are 98.89 pct identical to
residues 1 to 451 of 451 from Escherichia coli K-12 :
B4119"
/codon_start=1
/transl_table=11
/product="alpha-galactosidase"
/protein_id="AAP18670.1"
/db_xref="GI:30042948"
/translation="MMSAPKTIIGAGSTIFVNILGDFVHFREALKTAHIALMDIDPT
RLEESHIVRKUMDSAGASKITCTHQKEALDADVFVAFQIGVEPCTVDFVC
KRHLQQTADTLPGGIMRALRTIPLHWQICENNTTEVCPDATMLNYNMMNTWAM
YARYPIKQVGLCHSVQGAELARDLNDIPATLYRCAGINHWAFYLELERKTADGS
YVNLYPELLAAEAGQAPKNTIHNTRQNIIVRYEMFKLIGVFTSESEHFAEYTPWF
IKLGRDLTERYKVPIDEPYKRCVQOLANWHKELEYKNASRIDIKPSREVAFTPMA
INTGEPSEVLYGNVNDGLDNLPOGCCVEVACLVANGIQPTKVTLPESHIAALMQTN
INVQTLTLEALTENRDRVYHAMMDPHTAAVLGIDEIYALVDDLLIAAHGDLFSLWH
R"
7755..9164
/gene="melB"
/locus_tag="S3627"
/locus_tag="(5094..6002)
/complement(5094..6002)
/gene="melB"

```

```

/locus_tag="S3627"
/function="transport; Transport of small molecules:
Carbohydrates, organic acids, alcohols"
/notes="residues 1 to 469 of 469 are 96.37 pct identical to
residues 1 to 469 of 469 from Escherichia coli K-12 :
B4120"
/codon_start=1
/transl_table=11
/product="melibiose permease II"
/protein_id="AAP18671.1"
/db_xref="GI:30042949"
/translation="MTTKLSYGFAGKQFAIGIVMYMLYYTVDVVLGVLGVLGTLF
LVAIRWADINPIMGTINATRSWGFKFWILGITLANSVILFLFSAHLEFGTQI
FVCVTVILMGMTVITMDIPWSLVPTITLDRERQOLVPPFPFASLAGFVAGTVL
IPVALLYKNDQSLGRLGPGFQMTLVLI AFFIVSTIITLRNVHEVFSDDNQSAGSHLTKA
IVALLYKNDQSLGRLGPGFQMTLVLI AFFIVSTIITLRNVHEVFSDDNQSAGSHLTKA
VTLVFPRLVKSLSRRLIMAGASLIPVSCGVLMLMAMSHYNNVLLIVIGILLNVGT
AFNFWLQVIMADVTDYGEYKLVHRCESIAYSQTVMVVKGSGAPAAFFIAVLMIGY
VFNVEQTAALLGQFIMIALPTLFFMVTILYFRFYRLNGDTRLRIQIHLDDKRYK
PPEPVHADIPVGAUSDVKA"
9250..9283
/notes="REP (repetitive extragenic palindromic) element,
contains 1 REP sequence"
/complement(9303..9932)
/gene="yjdF"
/locus_tag="S3628"
/complement(9303..9932)
/gene="yjdF"
/locus_tag="S3628"
/notes="residues 1 to 209 of 209 are 83.25 pct identical to
residues 1 to 209 of 209 from Escherichia coli K-12 :
B4121"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP18672.1"
/db_xref="GI:30042950"
/translation="WTRFLKPLIINTGALALTLLIYTGISANDKLTWLEVTPIIV
VPLLATARRPLTLTLTLFFHAILMVGQGYIYAKVPVGVFEVQWLGSLSRNPDK
Query Match 93.9%; Score 577.6; DB 1; Length 290976;
Best Local Similarity 98.1%; Pred. No. 4.7e-173;
Matches 606; Conservative 0; Mismatches 9; Indels 3; Gaps 2;
QY 1 ATGACCTTAGAATGGTGGTGGCTTACCTGCTGATCATCGATCATTTTAACGCTGTCGCCA 60
DB 251870 ATGACCTTAGAATGGTGGTGGCTTACCTGCTGATCATCGATCATTTTAACGCTGTCGCCA 251929
QY 61 GGCTCTGTGGTCAATCAACACTATGACCACTCGCTCAACACCGGTTAT--CCGCGCCGGTG 118
DB 251930 GGCTCTGTGGTCAATCAACACTATGACCACTCGCTCAACACCGGTTATCGCGCGCGGTG 251989
QY 119 GCGTCTATGTGGCTTCAGACCGGACTGCGGATTCATATTGTGCTGTTGCGTGGG 178
DB 251990 GCGTCTATGTGGCTTCAGACCGGACTGCGGATTCATATTGTGCTGTTGCGTGGG 252049
QY 179 TT--GGGAGCGCTATTTTCCCGCTCAGTTCATTCGTTTGAAGTGTGGAAGTGGCGAGCGCG 237
DB 252050 TTGGGAGCGCTATTTTCCCGCTCAGTTCATTCGTTTGAAGTGTGGAAGTGGCGAGCGCG 252109
QY 238 GCTTACTTGAATTTGGCTGGGAATCAGCAGTGGCGCGCGCGCTGTGCAATTCACCTTAAA 297
DB 252110 GCTTACTTGAATTTGGCTGGGAATCAGCAGTGGCGCGCGCGCTGTGCAATTCACCTTAAA 252169
QY 298 TCGCTGGCGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCGAGTTTTTGTGAATCTC 357
DB 252170 TCGCTGGCGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCGAGTTTTTGTGAATCTC 252229
QY 358 ACCAATCCCAAAAGTATTGTGTTCTGGCGCGGCTATTTCGCGAATTCATCATGCCGCCAA 417
DB 252230 ACCAATCCCAAAAGTATTGTGTTCTGGCGCGGCTATTTCGCGAATTCATCATGCCGCCAA 252289
QY 418 CAGCGCGCAACTGATGACGTATATCGTGTCTCGCGCGTCCACCACTATTGTGGTCGATATTAT 477

```

```

Db 252290 CAGCCGCACTGATGAGTATATCGTCTCGGCTCACCACTATTGTGTGCTGATATTATT 252349
Qy 478 GTGATGATCGGTTACGCCACCCCTTGCTCAACGATTTGCTCTATGATTAAGGACCAAG 537
Db 252350 GTGATGATCGGTTACGCCACCCCTTGCTCAACGATTTGCTCTATGATTAAGGACCAAG 252409
Qy 538 CAGATGAAGCGCTGATGAATAGATTTTCGGCTCGTGTGTTTATGCTGCTGGGAGCGCTGTTA 597
Db 252410 CAGATGAAGCGCTGATGAATAGATTTTCGGCTCGTGTGTTTATGCTGCTGGGAGCGCTGTTA 252469
Qy 598 GCATCGCGGAGGCGATCGG 615
Db 252470 GCATCGCGGAGGCGATCGG 252487

RESULT 9
AP002567/c
LOCUS      318703 bp      DNA      linear      BCT 07-MAR-2001
DEFINITION      Escherichia coli O157:H7 DNA, complete genome, section 18/20.
ACCESSION      AP002567 BA000007
VERSION        AP002567.1 GI:13364198
KEYWORDS
SOURCE
ORGANISM      Escherichia coli O157:H7
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS      Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
Yamanoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL
MEDLINE      20198780
PUBMED       10734605
REFERENCE
AUTHORS      Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
Hayashi,T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL
MEDLINE      20557356
PUBMED       11108008
REFERENCE
AUTHORS      Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
Yamanoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
Shinagawa,H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
JOURNAL
MEDLINE      20564182
PUBMED       11111050
REFERENCE
AUTHORS      Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
Shinagawa,H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
JOURNAL
MEDLINE      21156231
PUBMED       11258796
REFERENCE
AUTHORS      Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
Hayashi,T.

```

```

TITLE      Direct Submission
JOURNAL     Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES
source
1..318703
/organism="Escherichia coli O157:H7"
/mol_type="genomic DNA"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
103..648
/gene="ECs4723"
103..648
/gene="ECs4723"
/feature="similar to WECD ECOLI gi|1790225 percent identity
98 in 181 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38146.1"
/db_xref="GI:13364199"
/translation="MQAKIASNTGELDALQQLGFSILVEGEVDLALPNNASDGAVV
AQETDIPALROLAAGAAFAQSPFRAPWYAPDASSRFYAQWIENAVRGTFDQCILIRAA
SGDIRGVYSLRELNATDARIGLLAGAGAEMLQALNWAYARGKTLRLVATQMGNTA
ALKRYIOSGANVESTAYLWR"
553..1783
/gene="ECs4724"
553..1783
/gene="ECs4724"
/feature="similar to WECE ECOLI gi|2367285 percent identity
99 in 376 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative regulator"
/protein_id="BAB38147.1"
/db_xref="GI:13364200"
/translation="WIPNPAPPVVGTVLDYMQSAMSGSKLGDGGFTTRCQQLQORF
GSAKVLTPSTASLEMANLLDIQRGEVIMPSYTFVSTANAFVLGAKIVFDVVP
YKMDITLIEAATDKTRIVPVHYAGVACEMDTINALAKNLFVVEAAQGVMT
DTGRALGTIGHGCFSTHETKNTYTAGGEGATLINDKALIERAIIIEKGTNRSGQFR
GVDRKYTRDITGSSYLMQAAYLMAQLEADRINQORLALMNYDVALAPLAKAGR
IELPSIDGCLQNAHMFYIKRLDITGDSALINFLKEAEINAVPHYIPLHGPCAGERFG
EFGEDRYTTKESERLLRLPLFYNLSPVNQRTVIATLLNYS"
1785..3035
/gene="ECs4725"
1785..3035
/gene="ECs4725"
/feature="similar to WXKE ECOLI gi|1790227 percent identity
98 in 416 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative cytochrome"
/protein_id="BAB38148.1"
/db_xref="GI:13364201"
/translation="MSLAKASLWTAASLTVKIGAGLVGKLAIVSQFAGLGAANFR
QLITVLGVLGAGIFNGVTKYVAQYHNDNPQOLRRVVGTSSAVLGFSTLMAVFLVAA
APISLIGFNHGYGLRVLVALVQMGIAWGNLLALMKGFDAAGNALSILVGLIGV
LAYVSVRLGGVEGALLGLALIPALVVI PAVMILKGAIPLSVLKPSWNGLAGOLS
KFTMLTITSVTLPVAYIMRKLAAQYSDWGVGICQVSSISDAVLOFITASEFSYL
LPTLSRLTKEDITREPVKSLKFLVPAPAAASFTVLLLRDFALWLLSNKFTAWROLF
AWQLVDVLYKGVAYFGVLYVIKASLRFYILAQVSTFLLMVFAMHLIPTHGALGAQ
AYMATVIVYFSLCCGFLWRRRA"
3032..4111
/gene="ECs4726"
3032..4111
/gene="ECs4726"

```

```
/notes="N-terminal part (1-62 in 359 aa) is similar to
YIFM1 ECOLI gi|2367286 percent identity 96 in 62 aa,
C-terminal part (156-359 in 359 aa) is similar to
YIFM2 ECOLI gi|2367287 percent identity 96 in 204 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38149.1"
/db_xref="GI:13364202"
/translation="MTVLHVLGSDIPHNRTVLRFNDALAAATSEHAREBMVVGKDD
GLSDSCPALSVDPFPCKSLAEAVLAKANKAQQRPFPFGQNPFTLWLALLSGIKXS
LYFTHWAGDLVLSGLLYKLYFLRLRLAQRVCVFATRGDLSFFAKTHPKVRGSL
LFTPTMDASLNTMANDROREKMTLVGNSGDRSNEHIAALRAVHOOGFTVTVVFP
MPPHNEYAIEVRQAGLELSEENLVLSKLEFDAYLLTRQCALGFYIFARQG
IGTCLLIQAGIPCVLRENPFWDMTQEHLEFLFTDDLNEIVREARQLASVDKN
TIAFFSPNYLQWQORALAAAGEVA"
/108..5460
/gene="ECe4727"
/108..5460
/gene="ECe4727"
/notes="similar to WECF_ECOLI gi|2367288 percent identity
99 in 450 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="UDP-Fuc4Nac-lipidII transferase"
/protein_id="BAB38150.1"
/db_xref="GI:13364203"
/translation="MSLLQFSGLFVYVWLLCTFIATLWFRFRVRNFNVFSLFLF
LTFPPFPPTSLVLPFDVGVAPPITLQALLSAGCFVAVVYVYKRLRKRVDVPR
RLPTVMNVTNLTWILMGIALVSGIFPMHNGFLFLRLNSYQISFSEVSGVALKR
FYEFYFPAWLAVLRQDSKAMFLVAVFAGLLTYMIVSGTRANIIAPAIPLFVG
IIFGMSLAWMLAAAGLVIGVGFMLALKRYGNVSGDEAFYFLYLRDTPSPENLA
LLQNDINDFQGLAEIVDFVFPFSLMPGRSPNSANSYFTFWEVLNHSGLAS
PLTIGSLVWVGGLFPLGAIYVVGIIKWFDMLYELGNRTNRYKAAILHSFCGALF
NMTVLAREGLDSFVSFWFFIVVFGACLMIAXLWLPESAGLHKTKSLRQVEG
"
5463..6203
/gene="ECe4728"
5463..6203
/gene="ECe4728"
/notes="similar to WECG_ECOLI gi|2367289 percent identity
98 in 246 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="probable UDP-N-acetyl-D-mannosaminuronic acid
transferase"
/protein_id="BAB38151.1"
/db_xref="GI:13364204"
/translation="MNNTTAPTYTLRGLQIGMRDMQHALDYLFADGQLKQGLTVAI
NAEKMTIENAEVRELIINAAEFKYADGIVSVRSVKYPOQASRVAGADLWELMA
RACKGTTVFLVGKPELVLAQTEAKLRNOMNVIIVSGODGYFKPEQCALPERIHASG
AQIVTVAMGSPKQIEIMRDCRLVHPDALVMVGVTYDFTGHVKEAPKIMQTLGLEWL
YRLLSQPSRIKQLRLRLIRHYGNL"
6394..7779
/gene="ECe4729"
6394..7779
/gene="ECe4729"
/notes="similar to YIFK_ECOLI gi|2367290 percent identity
98 in 461 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative amino acid/amine transport protein"
/protein_id="BAB38152.1"
/db_xref="GI:13364205"
/translation="MADNKPQLQRLQEAHIELIAGTIGVLGFMGAASLTKWAGPS
VLAYIILAGLVFPIRNSGEMWLEFVPTGSAVVAHYMSFPFGYLTASYWPMWA
VGISETATGVVQVQFPFPMAGWIPALIAVALVALANLAARVLGGEIFFPAMIKVIT
IIVNIVIGLVVIFFGFGGSGIGFSNLTUEHGFFPAGGWNKGFLTALCIVVASVYGVEL
```

```
Query Match 92.9%; Score 571.2; DB 1; Length 318703;
Best Local Similarity 97.4%; Pred. No. 5.4e-171;
Matches 602; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

QY 1 ATGACCTTAGAATGTTGGTTTGCCTACCTGCTGACATCGATCATTTTAACGCTTCGCCA 60
DB 37470 ATGACCTTAGAATGTTGGTTTGCCTACCTGCTGACATCATTTTAACGCTTCGCCA 37411
QY 61 GGCCTGTGGTGCACATCAACACTATGACACACCTCGCTCAACACCGGTAT - -CCGCGCCGGTG 118
DB 37410 GGCCTGTGGTGCACATCAACACTATGACACACCTCGCTCAACACCGGTATTCGCGCGCGGTG 37351
QY 119 GGCCTATTTGCTGGCTTCAGACCGGACTGGGATTCATATTGCTGGTGGTGGCTGGGG 178
DB 37350 GGCCTATTTGCTGGCTTCAGACCGGACTGGGATTCATATTGCTGGTGGTGGCTGGGG 37291
QY 179 TT -GGGACGCTATTTTCCCGCTCAGTGATTCGCTTTGAAGTGTTCGAGTGGCGGAGGGCG 237
DB 37290 TTGGGAGCGCTATTTTCCCGCTCAGTGATTCGCTTTGAAGTGTTCGAGTGGCGGAGGGCG 37231
QY 238 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAA 297
DB 37230 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGACCTTAAA 37171
QY 298 TCGCTGGCTCTACTCAATCGCTGCGACATTTGTTCCAGCGCGCAGTTTTTGTCAATCTC 357
DB 37170 TCGCTGGCTCTACTCAATCGCTGCGACATTTGTTCCAGCGCGCAGTTTTTGTCAATCTC 37111
QY 358 ACAATPCCAAAGATTTGTGTTCTGGCGCGCTATTTCCGCAATTCATCATGCCGCAA 417
DB 37110 ACAATPCCAAAGATTTGTGTTCTGGCGCGCTATTTCCGCAATTCATCATGCCGCAA 37051
QY 418 CAGCGCGCAACTGATCGCATATATCGTCTCGCGCTCACCACATTTTGTGGTGCATTTATT 477
DB 37050 CAGCGCGCAACTGATCGCATATATCGTCTCGCGCTCACCACATTTTGTGGTGCATTTATT 36991
QY 478 GTGATGATCGTTTACGCCACCTTCTCTCAACGGATTCGCTATCGATTAAAGGACCAAAG 537
DB 36990 GTGATGATCGTTTACGCCACCTTCTCTCAACGGATTCGCTATCGATTAAAGGACCAAAG 36931
QY 538 CAGATGAAGCGCTCAATAAGATTTTCGGCTCGCTGTTTATGCTGGTGGGAGCGCTGTTA 597
DB 36930 CAGATGAAGCGCTCAATAAGATTTTCGGCTCGCTGTTTATGCTGGTGGGAGCGCTGTTA 36871
QY 598 GCATCGGCGAGGCATGCG 615
DB 36870 GCATCGGCGAGGCATGCG 36853

RESULT 10
AE016769/c
LOCUS
DEFINITION
Escherichia coli CFT073 section 15 of 18 of the complete genome.
ACCESSION
AE016769 AE014075
VERSION
AE016769.1 GI:26110701
KEYWORDS
SOURCE
Escherichia coli CFT073
ORGANISM
Escherichia coli CFT073
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 301566)
Welch,S.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
PUBMED
12471157
2 (bases 1 to 301566)
Welch,S.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
```

Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,P.R.
Direct Submission

TITLE
JOURNAL
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

```

source
1..301566
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(72..1100)
/gene="waaw"
/locus_tag="c4450"
complement(72..1100)
/gene="waaw"
/locus_tag="c4450"
/notes="Residues 1 to 341 of 342 are 98.24 pct identical to
residues 1 to 341 of 341 from GenPept.129:
>gb|AAC69673.1| (AF015974) UDP-galactose (galactosyl) LPS
alpha1,2-galactosyltransferase Waaw [Escherichia coli]"
/codon_start=1
/transl_table=11
/product="UDP-galactose (galactosyl) LPS
alpha1,2-galactosyltransferase"
/protein_id="AAN82886.1"
/db_xref="GI:26110702"
/translation="MDLAESITEVAVSGEITANTDRVLNIAYGIDRNFPGAAMVSMOS
VHNPDLAVKHLFTDYIDEDYLRQVNAFTSKANVEVRIYKVSSAFIDIFPSLKQW
SVATFRLVAFQYSETIENLIDADYICKSGLAGLIDNFDGKFAAVIKDPFQW
EKPAKLAIEGPNFYGVYVQLQLEAKWKNFNMKAIAMLASDPQHTYKICLDODI
LNLFFSHCIFISGDCDFYGVYVQLQLEAKWKNFNMKAIAMLASDPQHTYKICLDODI
NYPCKQVNEAYQVSCWMDVAFIPATNEKQYQVYQAHKXGDTFNAPFVIFKPLNK
YRKLPQW"
complement(1126..1818)
/gene="rfaj"
/locus_tag="c4451"
complement(1126..1818)
/gene="rfaj"
/locus_tag="c4451"
/function="enzyme; Macromolecule metabolism;
Lipopolysaccharide"
/notes="Escherichia coli K-12 ortholog: b3625; Escherichia
coli O157:H7 ortholog: z5052"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide core biosynthesis protein
rfaj"
/protein_id="AAN82887.1"
/db_xref="GI:26110703"
/translation="MITSIRVGFYKNDNKNYKEIFDEILAYNFKTVKVLNIDD
TKVSLIDTKYGVYFVFKPKRNERFKSGVKGQYQNLIVETDRVSGALFTFND
FYFLAERKIFNYSVFIILIEYVEGVLENDMPIIPEDVKARIKASMEKHLALNLSGD
PRHGNFVSKDGVRIIDLSGKSCAERKARDRLAMERHLGTANEIKDYGVYVYRTK
LRKFLKKGKA"
complement(1828..2823)
/gene="rfaj"
/locus_tag="c4452"
complement(1828..2823)
/gene="rfaj"
/locus_tag="c4452"
/BC_number="2.4.1.58"
/function="enzyme; Surface polysaccharides and antigens"
/notes="Escherichia coli K-12 ortholog: b3626; Escherichia
coli O157:H7 ortholog: z5051"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide 1,2-glucosyltransferase"
/protein_id="AAN82888.1"
/db_xref="GI:26110704"
/translation="WNEFIKRFVSLADNKENVPELVNSYGIDKNFLYAGVYSISV
LINNSDINVFVFTDYDDYKLSFNETAKQFNSTSIIVLIDPKFADLPSTQFWSY
ATYFVLSFEYSIESITLLLDADVCKSGSKLPTKIIIFKDEFAVVPNDSTQAC
AKRLNIPGMNGRYFNAGVIYNLKKWHEANLTPVLLTLLRGTKYGLKYLDQDALNI

```

```

AFNMNVIYAXDFTIYTKNELHGRHRKYQOITTKTVLHVITGITKPMHWSWAGYP
SASVFNIAREQSPMKYPLKEARTVAEMQKQKXELFAHGEVIKITSILYKLLKK"
complement(2840..3856)
/gene="rfaj"
/locus_tag="c4453"
complement(2840..3856)
/gene="rfaj"
/locus_tag="c4453"
/BC_number="2.4.1.44"
/function="enzyme; Surface polysaccharides and antigens"
/notes="Escherichia coli K-12 ortholog: b3627; Escherichia
coli O157:H7 ortholog: z5053"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide 1,3-galactosyltransferase"
/protein_id="AAN82889.1"
/db_xref="GI:26110705"
/translation="MSAHYFNPOEWINKTIIPDERPAVSVASSPHVAYGIDQNFPGC
GVSTSVLHNDVSFVHFVIDIPADQRLSOLAKSVHTCIQHLYNCRKALP
TKNWSIAYFRFVIADYFIDQDKILYLDADIACQGNLKLPIITMOANVAVVTER
DANWMSRAQSLQCNLEKGFNSGVILLINTLAWAESVSAKMSMLADKAVVSLTY
MDQDILNLLIGKVKFIDAKYNTQFSLNYELKKSFCIPINDETVLIIHYVGTPEWHYV
AGYPSAQPFIKGEASPMKNEPLMRPMNSNYARYCAKHNPKQNKPIINGIMNYIYFYL
KIHK"
complement(3872..4678)
/gene="rfaj"
/locus_tag="c4454"
complement(3872..4678)
/gene="rfaj"
/locus_tag="c4454"
/function="enzyme; Macromolecule metabolism;
Lipopolysaccharide"
/notes="Escherichia coli K-12 ortholog: b3630; Escherichia
coli O157:H7 ortholog: z5054"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide core biosynthesis protein
rfaj"
/protein_id="AAN82890.1"
/db_xref="GI:26110706"
/translation="MVMMVLEKEPFAIWRGKOPFERVKTIQGEVFPRELETRTLRPE
MAGSYFLKWRHGGTTLEIKNLSLPMVIGADREWNAIHLRLDVGVDVTVYGAFFE
KGNPLRTSFITITEDTPTISLEDYCADWATNPPDVRVKRMLKRVATVYRDMHAAG
INHRDVCYICHELHPFSGKEELKISVIDLHRAQLRTRVPRWRDKDLGLGYFSSMN
IGLTORDIWRPMKVYFSAPLKDLKQSGLLSQAEAKATKIRETRKSL"
complement(4662..5786)
/gene="rfaj"
/locus_tag="c4455"
complement(4662..5786)
/gene="rfaj"
/locus_tag="c4455"
/function="enzyme; Macromolecule metabolism;
Lipopolysaccharide"
/notes="Glucosyltransferase 1; Escherichia coli K-12
ortholog: b3631; Escherichia coli O157:H7 ortholog: z5055"
/codon_start=1
/transl_table=11
/product="Lipopolysaccharide core biosynthesis protein
rfaj"
/protein_id="AAN82891.1"
/db_xref="GI:26110707"
/translation="MIVAFCLYKFPFGQLQDRPMIAQTVAAARGHVRVYTQSWEGE
CPDFEILIKVPSKSTNHRNABYFAWVKHLREHPVDVRVVGKMPGLDVIYAADVC
YAEVAKGEKGFYELTRYHAAFERATFEQKQTOLLMLTDKQADFOKHYOTAE
RPHLPQIYPRKYQQPANSREIFPKNGITEQQYLLQVSGDFTKRGVDSREIAL
ASLPDSLRHNTLLYVQDKPRKFAEKRGVRSNVHFFSGRNDVSELMAAADLLIH
PAYOAAIGVLELATAGLPVLTAVCGYAHYIVDANCGEAIAEPFRQETLINEILKA
LTQSSLRQAWEANARHYADTDLYSLPEKAAIDIITGDLG"
complement(5783..6841)
/gene="rfaj"
/locus_tag="c4456"
complement(5783..6841)
/gene="rfaj"

```


QY	538	CAGATGAAGGCGCTGAATAGATTTCCTGCTCGTTTATGCTTGTGGAGCGCTGTTA	597
Db	280654	CAGATGAAGGCGCTGAATAGATTTCCTGCTCGTTTATGCTTGTGGAGCGCTGTTA	280595
QY	598	GCATCGGCGAGGCATGCG	615
Db	280594	GCATCGGCGAGGCATGCG	280577
RESULT 11			
AE005613/c			
LOCUS		15992 bp	linear
DEFINITION		Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 232 of 290.	BCT 21-MAR-2001
ACCESSION	AE005613	AE005174	
VERSION	AE005613.1	GI:12518685	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PMID			
PUBMED			

```
CDS
1206..2156
/gene="corA"
/function="transport; Transport of small molecules;
Cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical
to residues 1 to 316 of 316 from Escherichia coli K-12
Strain MG1655; B3816"
/codon_start=1
/transl_table=11
/product="Mg2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="MISAFQLENNRLTLEVESQPLVNAVWIDLVPDDDERLVOS
EQGSLATPELEDIEASARFEDDDGHHHSFFFEADHAGNSTVAFITDGRIF
TIRELELPRIYRMRQSNVDGNAYELLDDLFETKIQLEADEIENIYSDLEQLSR
VIMEHQGDEYALSTLAELDIEGWKRLCLMDTQALNPLVRKARLPGQLEQARE
IIRDIESLPHNESIPQKVFQMAAGFINIEQNRKIKIPSVSVVFLPPTLVASSY
GNRFEFPELKWSPGPGALIFMILAGLAPLYFKKNWL"
2157..4848
/misc_feature
/note="O-island #156; Region of the HDL933 chromosome not
homologous to E. coli K-12 MG1655"
/complement(2198..2689)
/gene="25334"
/complement(2198..2689)
/gene="25334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MKSLOKILICLLPCCYINLWLRPVQILYTVSDAGNSVFLV
DILPWTDSKINWYLNKHONEIKNOHPLPESGWHYVIDGNGFTDYKKYIEPYEDL
YCFPIIKENDNCITKYNLVINEPYRNTHTGINDFEYQITQENKIEQVFNPNFKK
DNF"
/complement(2686..3534)
/gene="25335"
/complement(2686..3534)
/gene="25335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNTLLPSINIPCTPETISLPDYSADDMQYDMVQDFLSLG
LSDIAKVDVRLKYHPEPGGSINVAFSASSGTSKISQRECDILFAEMKELAKNFS
FPGQKYLIELLIEHFRYNGSNFHSQQLNLSFEKINKYGYNSPIRIKECIENGIN
SPISGYQLIQSLIKTKLLSLRNKFNDFSPNGLIGIVHDISAQKISLSLSFKYA
IGWSATIHVAQDHGLDVTDIRKNTYSKYRFFRIWFLQHKDFAPKPFPTNFETIE
RIENYL"
3748..3867
/gene="25336"
3748..3867
/gene="25336"
/function="orf; Unknown function"
/note="Residues 1 to 39 of 39 are 100.00 pct identical to
residues 278 to 316 of 316 from GenPept 118 :
gi|6960250|gb|AA033440.1| (AF233324) S. typhimurium
magnesium transport protein (CORA) (SP:P31138) [Salmonella
typhimurium Lt2]"
/codon_start=1
/transl_table=11
/product="partial repeat of corA"
/protein_id="AAG59012.1"
/db_xref="GI:12518690"
/translation="MNFPEFMPLELKWSPGPGALIFMILAGLAPLYFKKNWL"
/complement(3892..4278)
/gene="25337"
/complement(3892..4278)
/gene="25337"
/function="orf; Unknown function"
/note="No significant matches"

misc_feature
1206..2156
/gene="corA"
/function="transport; Transport of small molecules;
Cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical
to residues 1 to 316 of 316 from Escherichia coli K-12
Strain MG1655; B3816"
/codon_start=1
/transl_table=11
/product="Mg2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="MISAFQLENNRLTLEVESQPLVNAVWIDLVPDDDERLVOS
EQGSLATPELEDIEASARFEDDDGHHHSFFFEADHAGNSTVAFITDGRIF
TIRELELPRIYRMRQSNVDGNAYELLDDLFETKIQLEADEIENIYSDLEQLSR
VIMEHQGDEYALSTLAELDIEGWKRLCLMDTQALNPLVRKARLPGQLEQARE
IIRDIESLPHNESIPQKVFQMAAGFINIEQNRKIKIPSVSVVFLPPTLVASSY
GNRFEFPELKWSPGPGALIFMILAGLAPLYFKKNWL"
2157..4848
/misc_feature
/note="O-island #156; Region of the HDL933 chromosome not
homologous to E. coli K-12 MG1655"
/complement(2198..2689)
/gene="25334"
/complement(2198..2689)
/gene="25334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MKSLOKILICLLPCCYINLWLRPVQILYTVSDAGNSVFLV
DILPWTDSKINWYLNKHONEIKNOHPLPESGWHYVIDGNGFTDYKKYIEPYEDL
YCFPIIKENDNCITKYNLVINEPYRNTHTGINDFEYQITQENKIEQVFNPNFKK
DNF"
/complement(2686..3534)
/gene="25335"
/complement(2686..3534)
/gene="25335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNTLLPSINIPCTPETISLPDYSADDMQYDMVQDFLSLG
LSDIAKVDVRLKYHPEPGGSINVAFSASSGTSKISQRECDILFAEMKELAKNFS
FPGQKYLIELLIEHFRYNGSNFHSQQLNLSFEKINKYGYNSPIRIKECIENGIN
SPISGYQLIQSLIKTKLLSLRNKFNDFSPNGLIGIVHDISAQKISLSLSFKYA
IGWSATIHVAQDHGLDVTDIRKNTYSKYRFFRIWFLQHKDFAPKPFPTNFETIE
RIENYL"
3748..3867
/gene="25336"
3748..3867
/gene="25336"
/function="orf; Unknown function"
/note="Residues 1 to 39 of 39 are 100.00 pct identical
to residues 278 to 316 of 316 from GenPept 118 :
gi|6960250|gb|AA033440.1| (AF233324) S. typhimurium
magnesium transport protein (CORA) (SP:P31138) [Salmonella
typhimurium Lt2]"
/codon_start=1
/transl_table=11
/product="partial repeat of corA"
/protein_id="AAG59012.1"
/db_xref="GI:12518690"
/translation="MNFPEFMPLELKWSPGPGALIFMILAGLAPLYFKKNWL"
/complement(3892..4278)
/gene="25337"
/complement(3892..4278)
/gene="25337"
/function="orf; Unknown function"
/note="No significant matches"

gene
1206..2156
/gene="25339"
4483..4794
/gene="25339"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59014.1"
/db_xref="GI:12518692"
/translation="MDSIWENVTFGSCWYNPYSKSDVFKAWLLMWAITILVPLT
SLTTLNLRISIGISFGCOILGRYTMISLGNWSKYWGVLGFPVTVNLFLILIFPKK
R"
/complement(4841..5728)
/gene="rard"
/note="synonym: 25340"
/complement(4841..5728)
/gene="rard"
/function="orf; Unknown function"
/note="Residues 4 to 295 of 295 are 92.12 pct identical to
residues 8 to 299 of 300 from Escherichia coli K-12 Strain
MG1655; B3819"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59015.1"
/db_xref="GI:12518693"
/translation="MDAKTRQGVLLALAAAYFIWGIAPAYFKLIYVYPADEILTHRVI
WSFFFMVLSICRQMSYKLTIQPKIFMLAVSAVLIGNWLLFTFWVNNHMLEA
SLGYFINPLVNIVLGMIFLGERFRMQLAVILAIQGLVQLWTFGLSPIIAGLAFS
FARYGLVRKTAVERAOTGLIETMWLLPVAIYLFALADSSHTSMGQNPMSALLLIA
AGIYTVPLICFPAATRLRLSLTLGFGFYIGTIFMELLAVTFYGEKPGADKQVTFEII
WVALAIFVMDAIRYQRRKS"
/complement(5780..6265)
/gene="yig"
/note="synonym: 25341"
/complement(5780..6265)
/gene="yig"
/function="orf; Unknown function"
/note="Residues 1 to 161 of 161 are 100.00 pct identical
to residues 1 to 161 of 161 from Escherichia coli K-12
strain MG1655; B3820"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59016.1"
/db_xref="GI:12518694"
/translation="MKCKDDMSAVLTAEQALKLVGEMFYVHMPFNRLALGMELEYEKE
FAQLAFKNQPMVGNWQAQSIHGVTASALDVAGLVCGVSTLRHETISELQRQL
SRMGITDLRVLYLRPGRXERFTATSSLLRAGNKVAVARVELHNEQLYASATATVW
G"
6412..7281
/gene="pldA"
/note="synonym: 25342"
6412..7281
/gene="pldA"
/function="enzyme; Central intermediary metabolism:
Phosphorus compounds"

Query Match 90.08; Score 553.8; DB 1; Length 10592;
Best Local Similarity 97.2%; Pred. No. 1.2e-165;
Matches 584; Conservative 1; Mismatches 13; Indels 3; Gaps 2;

QY 18 GTTTGCCTACTGTCGATCGATCAATTTTAAAGCTGTGCCAGGCTCTGGTGCAATCAA 77
DB 10592 GTTTGCCTACTGTCGATCAATTTTAAAGCTGTGCCAGGCTCTGGTGCAATCAA 10533
```

* NOTE. This is a 'working draft' sequence. It currently
* consists of 144 conigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the conigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1159:	contig of 1159 bp in length
*		gap of	unknown length
*	1160	2363:	contig of 1204 bp in length
*		gap of	unknown length
*	2364	3604:	contig of 1241 bp in length
*		gap of	unknown length
*	3605	5188:	contig of 1584 bp in length
*		gap of	unknown length
*	5189	6372:	contig of 1184 bp in length
*		gap of	unknown length
*	6373	7737:	contig of 1365 bp in length
*		gap of	unknown length
*	7738	9783:	contig of 2046 bp in length
*		gap of	unknown length
*	9784	11046:	contig of 1263 bp in length
*		gap of	unknown length
*	11047	12223:	contig of 1177 bp in length
*		gap of	unknown length
*	12224	13492:	contig of 1269 bp in length
*		gap of	unknown length
*	13493	14669:	contig of 1177 bp in length
*		gap of	unknown length
*	14670	15886:	contig of 1217 bp in length
*		gap of	unknown length
*	15887	17089:	contig of 1203 bp in length
*		gap of	unknown length
*	17090	18308:	contig of 1219 bp in length
*		gap of	unknown length
*	18309	19518:	contig of 1210 bp in length
*		gap of	unknown length
*	19519	20710:	contig of 1192 bp in length
*		gap of	unknown length
*	20711	21932:	contig of 1222 bp in length
*		gap of	unknown length
*	21933	23606:	contig of 1674 bp in length
*		gap of	unknown length
*	23607	25328:	contig of 1722 bp in length
*		gap of	unknown length
*	25329	26649:	contig of 1321 bp in length
*		gap of	unknown length
*	26650	27849:	contig of 1200 bp in length
*		gap of	unknown length
*	27850	29135:	contig of 1286 bp in length
*		gap of	unknown length
*	29136	30700:	contig of 1565 bp in length
*		gap of	unknown length
*	30701	31877:	contig of 1177 bp in length
*		gap of	unknown length
*	31878	33065:	contig of 1188 bp in length
*		gap of	unknown length
*	33066	34267:	contig of 1202 bp in length
*		gap of	unknown length
*	34268	35474:	contig of 1207 bp in length
*		gap of	unknown length
*	35475	36650:	contig of 1176 bp in length
*		gap of	unknown length
*	36651	37859:	contig of 1209 bp in length
*		gap of	unknown length
*	37860	39072:	contig of 1213 bp in length
*		gap of	unknown length
*	39073	40406:	contig of 1334 bp in length
*		gap of	unknown length
*	40407	41630:	contig of 1224 bp in length
*		gap of	unknown length
*	41631	42864:	contig of 1234 bp in length

```
* * gap of unknown length
* * 44200: contig of 1336 bp in length
* * gap of unknown length
* * 44201: contig of 1887 bp in length
* * gap of unknown length
* * 46088: contig of 2140 bp in length
* * gap of unknown length
* * 48228: contig of 1291 bp in length
* * gap of unknown length
* * 49519: contig of 1392 bp in length
* * gap of unknown length
* * 50911: contig of 1510 bp in length
* * gap of unknown length
* * 52421: contig of 1193 bp in length
* * gap of unknown length
* * 53614: contig of 1391 bp in length
* * gap of unknown length
* * 55005: contig of 1296 bp in length
* * gap of unknown length
* * 56301: contig of 1205 bp in length
* * gap of unknown length
* * 57506: contig of 1181 bp in length
* * gap of unknown length
* * 58687: contig of 1721 bp in length
* * gap of unknown length
* * 60408: contig of 1225 bp in length
* * gap of unknown length
* * 61633: contig of 1248 bp in length
* * gap of unknown length
* * 62881: contig of 1340 bp in length
* * gap of unknown length
* * 64221: contig of 1200 bp in length
* * gap of unknown length
* * 65421: contig of 1185 bp in length
* * gap of unknown length
* * 66605: contig of 2077 bp in length
* * gap of unknown length
* * 68683: contig of 2030 bp in length
* * gap of unknown length
* * 70713: contig of 1183 bp in length
* * gap of unknown length
* * 71895: contig of 1567 bp in length
* * gap of unknown length
* * 73463: contig of 1344 bp in length
* * gap of unknown length
* * 74807: contig of 1399 bp in length
* * gap of unknown length
* * 76205: contig of 1375 bp in length
* * gap of unknown length
* * 77580: contig of 1501 bp in length
* * gap of unknown length
* * 79081: contig of 1233 bp in length
* * gap of unknown length
* * 80315: contig of 1585 bp in length
* * gap of unknown length
* * 81900: contig of 1241 bp in length
* * gap of unknown length
* * 83141: contig of 1525 bp in length
* * gap of unknown length
* * 84665: contig of 1223 bp in length
* * gap of unknown length
* * 85888: contig of 1267 bp in length
* * gap of unknown length
* * 87155: contig of 1280 bp in length
* * gap of unknown length
* * 88435: contig of 1520 bp in length
* * gap of unknown length
* * 89955: contig of 1322 bp in length
* * gap of unknown length
* * 91277: contig of 1236 bp in length
* * gap of unknown length
* * 92513: contig of 1404 bp in length
* * gap of unknown length
* * 93917: gap of unknown length
```

```
* * 93918 95448: contig of 1531 bp in length
* * gap of unknown length
* * 95449 96800: contig of 1352 bp in length
* * gap of unknown length
* * 96801 97978: contig of 1178 bp in length
* * gap of unknown length
* * 97979 99179: contig of 1201 bp in length
* * gap of unknown length
* * 99180 100437: contig of 1258 bp in length
* * gap of unknown length
* * 100438 101827: contig of 1390 bp in length
* * gap of unknown length
* * 101828 103117: contig of 1290 bp in length
* * gap of unknown length
* * 103118 104550: contig of 1433 bp in length
* * gap of unknown length
* * 104551 105854: contig of 1304 bp in length
* * gap of unknown length
* * 105855 107523: contig of 1669 bp in length
* * gap of unknown length
* * 107524 109208: contig of 1685 bp in length
* * gap of unknown length
* * 109209 110430: contig of 1222 bp in length
* * gap of unknown length
* * 110431 111662: contig of 1232 bp in length

Query Match 75.9%; Score 467; DB 2; Length 212936;
Best Local Similarity 96.4%; Pred. No. 1.3e-137; Indels 12; Gaps 11;
Matches 595; Conservative 0; Mismatches 10;

Qy 1 ATGACCTTGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA 60
Db 175539 ATGACCTTGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAAACGCTGTCGCCA 175481

Qy 61 GGCCTGTGGCAATCAACTATGACCACTGACCTGCTGCAACCAACCGTTAT--CGGCGCGGTG 118
Db 175480 GGCCTGTGGCAATCAACTAT--ACCACTGCTGCAACCAACCGTTATCGGCGCGGTG 175422

Qy 119 GCCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTATTTGCTGCTGCTGGCGTGGGG 178
Db 175421 GCCTCTATTGCT--GGCTTCAGACCGGACTGGCGATTATTTGCTGCTGCTGGCGTGGGG 175363

Qy 179 TTGGGACGCTATTTTCCGCTCAGTGATTGGTTTGAAGTGTGAAAGTGGGAGCGCGGG 238
Db 175362 TTGGGACGCTATTTTCCGCTCAGTGATTGGCTTTGAAGTGTGAAAGTGGGAGCGCGGG 175304

Qy 239 CTTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCAATTGACCTTAAAT 298
Db 175303 CTTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCAATTGACCTTAAAT 175244

Qy 299 CGCTGCGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTCA 358
Db 175243 CGCTGCGCTCTACTCAATCGCTCGACATTT--TTCAGCGCGCAGTTTGTGTAATCTCA 175185

Qy 359 CCAATCCAAAAGTATGTTGTTTCTGGCGCGCTATTTCCGCAATTCATCATCGCGCAAC 418
Db 175184 CCAATCCAAAAGTATGTTG--TTCGCGCGCGCTATTTCCGCAATTCATCATCGCGCAAC 175126

Qy 419 AGCGGCAACTGATGCAATATATGCTCGCGCTCGCGCTCAACCACTATTGTGCTGATATTG 478
Db 175125 AGCGGCAACTG--TGCAATATATGCTCGCGCTCGCGCTCAACCACTATTGTGCTGATATTG 175067

Qy 479 TGAATGATCGGTTACCGCACCTTGTCTCAACGATGCTCTATGGATTAAGGACCAACAGC 538
Db 175066 T-ATGATCGGTTACCGCACCTTGTCTCAACGATGCTCTATGGATTAAGGACCAACAGC 175009

Qy 539 AGATGAAGCGCTGAATAAGATTTTCCGCTCGTGTGTTTATGCTGCTGGAGCGCTGTTAG 598
Db 175008 AGATGAAGCGCTGAATAAGATTTTCCGCTCGTGTGTTTATG--TGCTGGAGCGCTGTTAG 174950

Qy 599 CATCGCGGAGGATCGG 615
Db 174949 CATCGCGGAGGATCGG 174933
```

[illegible]

* 46418 47815: contig of 1398 bp in length
* gap of unknown length
* 47816 49310: contig of 1495 bp in length
* gap of unknown length
* 49311 50304: contig of 994 bp in length
* gap of unknown length
* 50305 50848: contig of 544 bp in length
* gap of unknown length
* 50849 51828: contig of 980 bp in length
* gap of unknown length
* 51829 52699: contig of 871 bp in length
* gap of unknown length
* 52700 53365: contig of 666 bp in length
* gap of unknown length
* 53366 54633: contig of 1268 bp in length
* gap of unknown length
* 54634 55470: contig of 837 bp in length
* gap of unknown length
* 55471 56358: contig of 888 bp in length
* gap of unknown length
* 56359 56934: contig of 576 bp in length
* gap of unknown length
* 56935 57672: contig of 738 bp in length
* gap of unknown length
* 57673 58319: contig of 647 bp in length
* gap of unknown length
* 58320 58529: contig of 210 bp in length
* gap of unknown length
* 58530 59083: contig of 554 bp in length
* gap of unknown length
* 59084 59906: contig of 823 bp in length
* gap of unknown length
* 59907 60826: contig of 920 bp in length
* gap of unknown length
* 60827 61084: contig of 258 bp in length
* gap of unknown length
* 61085 61762: contig of 678 bp in length
* gap of unknown length
* 61763 62638: contig of 876 bp in length
* gap of unknown length
* 62639 62713: contig of 75 bp in length
* gap of unknown length
* 62714 63593: contig of 880 bp in length
* gap of unknown length
* 63594 64039: contig of 446 bp in length
* gap of unknown length
* 64040 65076: contig of 1037 bp in length
* gap of unknown length
* 65077 65366: contig of 290 bp in length
* gap of unknown length
* 65367 65590: contig of 224 bp in length
* gap of unknown length
* 65591 66041: contig of 451 bp in length
* gap of unknown length
* 66042 67040: contig of 999 bp in length
* gap of unknown length
* 67041 67560: contig of 520 bp in length
* gap of unknown length

Query Match 73.1%; Score 449.8; DB 2; Length 256373;
Best Local Similarity 95.7%; Pred. No. 4.5e-132;
Matches 495; Conservative 0; Mismatches 17; Indels 5; Gaps 3;
QY 1 ATGACCTTAGAATGGTGGTTTCCTACCTGCTGATCGATCATATTAA--CGCTGTGCG 58
DB 77559 ATGACCTTAGAATGGTGGTTTCCTACCTGCTGATCGATCATATTAAAGCTCGTCGCC 77500
QY 59 CAGGCTCTGGTGAATCAACACTATGACACCTGCTCAACACGGTTAT--CGGCGCG 116
DB 77499 CAGGCTCTGGTGAATCAACACTATGACACCTGCTCAACACGGTTATCGCGCGCG 77440
QY 117 TGGCGTCTATTGCTGGGCTTCAGACGGACTGGCGATTCAATATTGCTGCTGCTGG 176

DB 77439 TGGCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCAATTTGTGCTGTTTGGCGTGG 77380
QY 177 GGTTC--GGACAGCTATTTCCCGCTCAGTGATGCGTTTGAAGTGTTCGAAGTGGCGAGCG 235
DB 77379 GGTTCGGGAGCAGCTATTTTCCCGCTCAGTGATGCGTTTGAAGTGTTCGAAGTGGCGAGCG 77320
QY 236 CGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGCAATTCACCTTA 295
DB 77319 CGGCTTACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTTGCAATTCACCTTA 77260
QY 296 AATCGCTGGCTTACTCAATTCGGGTGACATTTGTTCCAGCGCGCAGTTTGTGAATC 355
DB 77259 AATCGCTGGCTTACTCAATTCGGGTGACATTTGTTCCAGCGCGCAGTTTGTGAATC 77200
QY 356 TCACCAATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCGCG 415
DB 77199 TCACCAATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCGCG 77140
QY 416 AACAGCCGCAACTGATGAGTATATCGTGTCTGCGGCGTCAACACTATTGTGTCGATATTA 475
DB 77139 AACAGCCGCAACTGATGAGTATATCGTGTCTGCGGCGTCAACACTATTGTGTCGATATTA 77080
QY 476 TTGTGATGATCGTTTACGCCACCCCTTGCTCAACGGAT 512
DB 77079 TTGTGATGATCGTTTACGCCACCCCTTGCTCAACGGAT 77043

RESULT 14
AR352333 417 bp DNA linear PAT 17-AUG-2003
LOCUS AR352333
DEFINITION Sequence 274 from patent US 6589738.
ACCESSION AR352333
VERSION AR352333.1 GI:33757296
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 417)
Unclassified.
AUTHORS Forsyth,R.A., Ohlsen,K. and Zyskind,J.W.
TITLE Genes essential for microbial proliferation and antisense thereto
JOURNAL Patent: US 6589738-A 274 08-JUL-2003;
FEATURES
Location/Qualifiers
1..417
source
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 67.3%; Score 414; DB 6; Length 417;
Best Local Similarity 100.0%; Pred. No. 4.6e-121;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 202 GTGATTGGTTTGAAGTGTTCGAAGTGGCGGCGGCTTACTTGTGATTGGCTGGGAATC 261
DB 1 GTGATTGGTTTGAAGTGTTCGAAGTGGCGGCGGCTTACTTGTGATTGGCTGGGAATC 60
QY 262 CAGCAGTGGCGCGCGCTGTGTCATTTGACCTTAAATCGCTGGCCTCTACTCAATCGCGT 321
DB 61 CAGCAGTGGCGCGCGCTGTGTCATTTGACCTTAAATCGCTGGCCTCTACTCAATCGCGT 120
QY 322 CGACATTTGTTCCAGCGCGCAGTTTGTGATCTCAATCCCAAAAGTATTGTGTTT 381
DB 121 CGACATTTGTTCCAGCGCGCAGTTTGTGATCTCAATCCCAAAAGTATTGTGTTT 180
QY 382 CTGGCGGCGCTATTTCGCAATTCATCATCGCAACAGCGCGCAACTGATGCAATATATC 441
DB 181 CTGGCGGCGCTATTTCGCAATTCATCATCGCAACAGCGCGCAACTGATGCAATATATC 240
QY 442 GTGCTCGGCGTCAACACTATTGTGTCGATATTTTGTGATGATCGTTAGCCACCCCTT 501
DB 241 GTGCTCGGCGTCAACACTATTGTGTCGATATTTTGTGATGATCGTTAGCCACCCCTT 300
QY 502 GCTCAACCGATTGCTCTATGATTAAAGGCAAGCAGATGAAGGCGCTCAATAGATT 561

Db 301 GCTCAAGGATGCTCTATGATTAAAGACCAAGACAGATGAAGCGCTGAATAAGATT 360
AL627278 258050 bp DNA linear BCT 04-JUL-2003
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
complete chromosome, segment 14/20.
QY 562 TTGGGCTCGTTGTTTATGCTGGGAGCGCTGTAGCATCGGAGCAGCATGCG 615
|||||
Db 361 TTGGCTCGTTGTTTATGCTGGGAGCGCTGTAGCATCGGAGCAGCATGCG 414
|||||
RESULT 15
AL627278 258050 bp DNA linear BCT 04-JUL-2003
LOCUS
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
complete chromosome, segment 14/20.
ACCESSION AL627278 AL513382
VERSION AL627278.1 GI:16504263
SOURCE
ORGANISM Salmonella enterica subsp. enterica serovar Typhi
Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 258050)
REFERENCE
AUTHORS Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,
Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,
Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
Conerton, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,
Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Garra, P.,
Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrall, B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
NATURE 413 (6858), 848-852 (2001)
JOURNAL Nature 413 (6858), 848-852 (2001)
MEDLINE 21534947
PUBMED 11677608
REFERENCE
AUTHORS Parkhill, J.
TITLE Direct Submission
SUBMITTED (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
JOURNAL
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
URL, http://www.sanger.ac.uk/Projects/s_typhi/).
FEATURES
source
1. 258050
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
69. .1544
/gene="tolC"
/note="synonym: STY3364"
69. .1544
/genes="tolC"
/note="Similar to Salmonella enteritidis outer membrane
protein TolC precursor tolC SW:TOLC_SALEN (Q54001) (491
aa) fasta scores: E(): 0, 99.0% id in 491 aa, and to
Escherichia coli outer membrane protein tolC precursor
tolC or mtcb or muka or ref1 SW:TOLC_ECOLI (P02930) (495
aa) fasta scores: E(): 0, 89.7% id in 495 aa
Orthologue of E. coli tolC (TOLC_ECOLI); Fasta hit to
TOLC_ECOLI (495 aa), 90% identity in 495 aa overlap"
/codon_start=1
/transl_table=11
/products="outer membrane protein TolC precursor"
/protein_id="CAD07712.1"
/db_xref="GI:16504264"
/db_xref="GOA:Q823N8"
/db_xref="SPTREMBL:Q823N8"
/translation="MQMKLLPILIGLSGFTSLQAEINLMQVYQQAISNPBLRKS
AADRAAFKINEARGSPLLPQLGLGADYTSNGRDANGINSNETSASLQITQLFDM

SKWRGLTQEKAAQIDVTYQTDQOTLILNTANAYFKVLNAIDVLSYTOACKAIVRQ
LDOTTFENVGLVAITVDONARQYDTVLANEVTRARNLNDAVEELRQVTHLYYPELA
SLNTEHFETDKKAVNALKEAENRNLILQARLSQDLAREQLRQADGHLPTLNLT
STGISDTSYSGSKTNSAQYDSDNMGQNKIGLNFSLPL:YQGMWNSQVKQYKVFVGAS
VLESAHRSVVQTVRRSFNNINASSINAYKQAVWSAQSSLDADWAGISVGRTIYVD
QIDATTLTDARQQLANARYTILINQLNIKYALGTLENEQDLLALNLTGLIGPIPTSPS
VAPETPEQDAADGYNASHAAPVQPTAARANSNNNGNFRH"
1757. .2428
/gene="STY3365"
1757. .2428
/gene="STY3365"
/note="Orthologue of E. coli ygiB (YGI_B_ECOLI); Fasta hit
to YGI_B_ECOLI (234 aa), 97% identity in 223 aa overlap.
Contains a possible N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="possible lipoprotein"
/protein_id="CAD07713.1"
/db_xref="GI:16504265"
/db_xref="SPTREMBL:Q8XG21"
/translation="MKRTKSIHHSFKRSARHLTPVALAVTAVFVLMACEKSDTV
SLYNADDCSAANFKSAECTTAYNNALKEARTAPKATREDCVAFSGGQQQAF
QAGMAPENQAQOQSGSGSFMPLMAGYMGRLMGAGFAQOQPLFSSKNPASPAYGY
TDAAGKNYGAAPQGRWTVPKTAMAPKPAUTTTVTRGGFGESVAKQSTMQSAAGTST
RSMGG"
1835. .1867
/gene="STY3365"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
2434. .3597
/gene="STY3366"
2434. .3597
/gene="STY3366"
/note="Fasta hit to YJFC_ECOLI (387 aa), 50% identity in
392 aa overlap
Orthologue of E. coli ygiC (YGI_C_ECOLI); Fasta hit to
YGI_C_ECOLI (386 aa), 94% identity in 386 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD07714.1"
/db_xref="GI:16504266"
/db_xref="SPTREMBL:Q8XFP6"
/translation="MERVITITERDWRDKATYGFNFHTMYGEFYWCEDAYKLTLAQ
VEKLDVTAELHQMLKVERVIAASDELMTKFRIPKHTGWVFRQSWTOQPSLYRLD
LAWDGIQPKLENNADPTSLYEAFQFWLQEDQINAGNLPEGSQDFNSLOEKLT
PFAELREYQYQGLLHLCRTDVEDRGTIOYLDQCAAEATIEFLYDIDIGKEKQ
FTDQDQVIANFLKLYPWFELRMSFTKLEDAGRWLEPAWKSIIENKILPLIWM
FDFENLLPAYFADEHPFDKYYVKTFISREGANWSIIENKIESVEGPGYGEEM
VQFYPLPKFGDSTLIGSWLINDQAPAGIGIREDRALITQDLSRFYPHIFVEG"
complement (3659. .4489)
/gene="STY3367"
/note="complement (3659. .4489)
/gene="STY3367"
/note="Orthologue of E. coli ygiD (YGI_D_ECOLI); Fasta hit
to YGI_D_ECOLI (271 aa), 87% identity in 266 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD07715.1"
/db_xref="GI:16504267"
/db_xref="GOA:Q823N7"
/db_xref="SPTREMBL:Q823N7"
/translation="WMVMTKLSCKDNIMSLTQMPALFLGHGSPMNVLDDNDYTRAWR
RLGEALPRPQAIYVVSAAHYTRGTGVTAMERPOTLHDFGGFPQALYDMHYPAQSPAL
AQRLLVELLAPVVALDKEAWGFDHGSWGLIKMYPNADIPMQLSVSDSTPAAHFEM
KRKLATLDEGVMLVAGNVVHNLTVRWHDGNIPIYPWAAAFNDYFVXANLTWQCPVEQ
HPLVNYLQHEGALSNPTEPHEFLPLLVVLGAWDGKEPITIPVDGIEMGSI SMLSVQVG
"
4588. .5361
/gene="STY3368"
4588. .5361
/gene="STY3368"

/note="Orthologue of E. coli ygiE (YGIIE_ECOLI); Fasta hit to ygiE_ECOLI (257 aa), 93% identity in 257 aa overlap. Contains multiple possible membrane spanning hydrophobic domains"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAD07715.1"
/db_xref="GI:16504268"
/db_xref="GOA:Q8XGR4"
/db_xref="SWISS-PROT:Q8XGR4"
/translation="MSVFLIILTLIAGAAATFICAFGLVGLGKPNRVLAFSLGPAAGIM LLSIMELPALDLEGGSPVLGYGMFTIGLGLYGLDLRLPHALPHDLVQKQOPLP GSKTALLILGLISLHNPSPGIATFTVASSNLELGFGLALAVHNPILPEGLARGPV VAATGKRTALPWAGISGMASILGLVLAHLISLVSPIYMAALMAVAGIMVALSVDELMPLEXIDENPNPSYGLGMSIMGLSLVILQIIG"
6004..7800
/gene="STY3370"
/note="synonym: asst"
6004..7800
/gene="STY3370"
/note="Similar to Klebsiella pneumoniae arylsulfate sulfotransferase asst TR:P97036 (EMBL:U32616) (599 aa) fasta scores: E(): 0, 86.8% id in 598 aa"
/codon_start=1
/transl_table=11
/product="probable arylsulfate sulfotransferase"
/protein_id="CAD07717.1"
/db_xref="GI:16504269"
/db_xref="GOA:Q8Z3N6"
/db_xref="SPTREMBL:Q8Z3N6"
/translation="MFDQYRTIILAGAVALTGSLTAASTFAAGFQPAQAGKIGAVV DPNAGPLTALVELDSHIIISDVKTIVHGKGEKVPVTVYTGKESLEYTDGIPIFGLYQ KPAVNTVYKENGAMKDDVYVOTSAIVNHYMDNRISIDLOOTKVIKVAPEGDERLY LVNHTFTPOAGEFHMGKEDKNAGILDAGPAGLPDIAPYTFVVDTCGEYRWLD QDTFYNGHDMNKIEGVLGMIRETPRGFTTAVQGHWEYFDMGQILADHKLPRGLD ASHESIVNGTVLLVRGRKEDGHHVHTIRDQIIIEVDKSRVVDVNDLTKLIDP MRDALLDAGAVCNVVDLAHQAKLEPDPYGDALGVAGRNVHNSIAYDAK DDSIILSRHQGIVKIGRQKWLIVAPSKGNKQLASKLLKXPDHKGPLTCDENK KXDTDFDTYQHTKWLISLGGTTLVFNQDGRGLEQALPTKMYSRFYKIDKKGKT VQVWEYKGRGYDFYSPISTVSEYQKDRDTMFGFGSINLFDVGKPTVKLNEIDYK TKEVKEVDLSDXPNQHYBALLVHTQMEK"
7820..8491
/gene="dsbA"
/note="synonym: STY3371"
7820..8491
/gene="dsbA"
/note="Similar to Enterobacter amnigenus disulfide isomerase dsbA TR:Q9XDP1 (EMBL:AF012826) (222 aa) fasta scores: E(): 0, 90.1% id in 222 aa"
/codon_start=1
/transl_table=11
/product="disulfide isomerase"
/protein_id="CAD07718.1"
/db_xref="GI:16504270"
/db_xref="GOA:Q8Z3N5"
/db_xref="SPTREMBL:Q8Z3N5"
/translation="MSSKWTSLSPKSVLVTALVTPFAASAFTEGTDVNWLEKPIPNA DKTLKVFSPFCYKIDKAVTGVDKVDLVTTPHLETKGYQASVEFVAVL IAKDAAGISLFDKAKFAWYAYHDKKRWSDGKDPAAFIKTLGLDAGMSQA DFEALADPAVQETLEKWAAYDVAKIQGVPAVYVNGKYLITKNIKSIDSMALVRE LATKK"
7847..8488
misc_feature

Query Match 61.7%; Score 379.2; DB 1; Length 258050;
Best Local Similarity 78.0%; Pred. No. 2e-109;
Matches 482; Conservative 0; Mismatches 133; Indels 3; Gaps 2;
QY 1 ATGACCTTAGAATGGTGGTTGGCTTACTGCTGACATCGATCATTTTAAAGCTGTGCGCA 60
|||||
Db 247744 ATGACCTTAGAATGGTGGTTGGCTTACTGACCTCACTCTGTGTGAGTCTTTCTCG 247803
|||||
QY 61 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACCAACCGGTTATCGGCGC--GTG 118

Search completed: March 28, 2004, 16:45:00
Job time : 2577.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 12:14:43 ; Search time 306.314 Seconds
(without alignments)
8529.292 Million cell updates/sec

Title: US-09-847-392-1_COPY_557_1171

Perfect score: 615

Sequence: 1 atgaccttagaatggtggtt.....tagcatggcgagcatgcg 615

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 337363 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_29Jan04.*
- 1: Geneseqn1980s.*
 - 2: Geneseqn1990s.*
 - 3: Geneseqn2000s.*
 - 4: Geneseqn2001as.*
 - 5: Geneseqn2001bs.*
 - 6: Geneseqn2002s.*
 - 7: Geneseqn2003as.*
 - 8: Geneseqn2003bs.*
 - 9: Geneseqn2003cs.*
 - 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	615	100.0	1200	3	Aa294405 E. coli x
2	615	100.0	1231	3	Aa48442 E. coli L
3	568.2	92.4	622	5	Aa93395 DNA encod
4	486.2	79.1	2216	5	Aa89415 DNA encod
5	414	67.3	417	4	Aa84646 E. coli g
6	414	67.3	417	7	ACA18626 Prokaryot
7	379.2	61.7	621	7	ACA51728 Prokaryot
8	375	61.0	618	7	ACA36103 Prokaryot
9	357.4	58.1	618	7	ACA32157 Prokaryot
10	344.8	56.1	623	7	ACA48809 Prokaryot
11	274.8	44.7	621	7	ACA53970 Prokaryot
12	259	42.1	259	7	ACA13399 Prokaryot
13	248.6	40.4	1884	5	Aa89412 DNA encod
14	207.4	33.7	263	4	Aa84479 E. coli g
15	175	28.5	274	5	Aa89410 DNA encod
16	167.8	27.3	516	5	Aa93398 DNA encod
17	149.8	24.4	618	7	ACA52910 Prokaryot
18	141	22.9	1041	5	Aa93397 Prokaryot
19	112.4	18.3	630	7	ACA42760 Prokaryot
20	104	16.9	630	7	ACA45284 Prokaryot
21	100	16.3	100	7	ACD71734 E. coli K
22	100	16.3	100	7	ACD71736 E. coli K
23	100	16.3	100	7	ACD71735 E. coli K

24	86.4	14.0	633	7	ACA43829	Prokaryot
25	80	13.0	627	7	ACA27048	Prokaryot
26	75.4	12.3	663	8	ADA29490	DNA encod
27	64.6	10.5	543	7	ACA20819	Prokaryot
28	58.8	9.6	633	7	ACA23772	Prokaryot
29	42.4	6.9	536	9	ADB68842	Minority
30	39.8	6.5	987	6	Abk94661	G protein
31	39.8	6.5	987	9	ADC51928	Mouse TGR
32	39.8	6.5	1083	6	Abk94660	G protein
33	39.8	6.5	1083	9	ADC51926	Mouse TGR
34	37.2	6.0	633	8	ADA31395	DNA encod
35	36.8	6.0	1377	4	ABL28479	Drosophil
36	36.8	6.0	5313	4	ABL28478	Drosophil
37	36.6	6.0	2000	7	ADA71938	Rice gene
38	35.4	5.8	1028	4	BAF79862	Soybean c
39	34.6	5.6	357	6	Abk77410	Bacillus
40	34.4	5.6	283	6	ABN16733	Human ORF
41	33.8	5.5	987	6	Abk94565	G protein
42	33.8	5.5	987	7	ABT33314	GR7 liga
43	33.8	5.5	987	8	ACC79865	Rat TGR26
44	33.8	5.5	987	9	ADC51916	Rat GPR8-
45	33.8	5.5	3292	9	ADB75566	Prostate

ALIGNMENTS

RESULT 1

AZ294405 ID AAZ94405 standard; DNA; 1200 BP.

XX AC AAZ94405; 18-JUL-2000 (first entry)

XX DE E. coli rhtB gene imparting homoserine resistance.

XX KW Homoserine resistance; rhtB gene; L-homoserine; L-alanine; L-isoleucine;

XX KW L-valine; L-threonine; ss.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers

XX FT CDS 557..1171

XX FT /*tag= a

XX FT /note= "a DNA sequence corresponding to the coding region (minus the stop codon) is specifically claimed in Claim 3"

XX EP994190-A2.

XX PD 19-APR-2000.

XX PF 20-SEP-1999; 99EP-00118581.

XX PR 13-OCT-1998; 98RU-00118425.

XX PA (AJIN) AJINOMOTO CO INC.

XX PI Livshits VA, Zakataeva NP, Aleoshin VV, Belareova AV; Tokhmakova IL;

XX DR WPI: 2000-273530/24.

XX PT P-PSDB; AAY79298.

XX PT Novel RhtB protein, useful for generation of L-homoserine resistance in Escherichia bacteria and large-scale production of e.g. L-homoserine and L-alanine.

XX PS Claim 3; Page 10-11; 14pp; English.

XX CC This is the DNA sequence the novel rhtB gene of Escherichia coli K-12. The rhtB gene maps at 86 min on the E. coli chromosome. The gene

CC participates in resistance to homoserine. Amplification of the gene
CC results in an improvement of the amino acid productivity of *E. coli*. The
CC invention provides: the RhtB protein (see AAY9298); DNA encoding the
CC RhtB, especially nucleotides 557-1171 of the present sequence; a
CC bacterium, especially of the genus *Escherichia*, in which L-homoserine
CC resistance is enhanced by amplifying the copy number or increasing the
CC expression rate of the *rhtB* DNA, the DNA being carried on a multicopy
CC vector or on a transposon; and a method for producing an amino acid by
CC cultivating the bacterium in a culture medium to produce and accumulate
CC the amino acid in the medium, from which it is recovered. The method is
CC used for the production of L-homoserine, L-alanine, L-isoleucine, L-
CC valine or L-threonine (all claimed)

SQ Sequence 1200 BP; 285 A; 309 C; 305 G; 301 T; 0 U; 0 Other;
Query Match 100.0%; Score 615; DB 3; Length 1200;
Best Local Similarity 100.0%; Pred. No. 2e-193;
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTAAACGTGCGCA 60
Db 557 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTAAACGTGCGCA 616
Qy 61 GGCTCTGGTCAATCAACACTATGACCACTGCTCAACACAGGTATCCGCCGGTGGC 120
Db 617 GGCTCTGGTCAATCAACACTATGACCACTGCTCAACACAGGTATCCGCCGGTGGC 676
Qy 121 GTCTATTGCTGGGCTTCAGACCGGACTGCGATTTCATATTGTGCTGGTGGCGTGG 180
Db 677 GTCTATTGCTGGGCTTCAGACCGGACTGCGATTTCATATTGTGCTGGTGGCGTGG 736
Qy 181 GGGACGCTATTTTCCCGCTCAGTTCGTTTGAAGTGTGAAGTGGGACGCGCGCT 240
Db 737 GGGACGCTATTTTCCCGCTCAGTTCGTTTGAAGTGTGAAGTGGGACGCGCGCT 796
Qy 241 TACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATTAATCG 300
Db 797 TACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATTAATCG 856
Qy 301 CTGGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 360
Db 857 CTGGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 916
Qy 361 AATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATCGCGCAACAG 420
Db 917 AATCCCAAAAGTATTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATCGCGCAACAG 976
Qy 421 CCGCAACTGATGCAATATATCGTCTCGGCTCAGCACTATTTGCTGCAATATTTCTG 480
Db 977 CCGCAACTGATGCAATATATCGTCTCGGCTCAGCACTATTTGCTGCAATATTTCTG 1036
Qy 481 ATGATCGGTTACGCCACCCCTTGCTCAACGATGCTCTATGGATTAAAGGACCAAGCAG 540
Db 1037 ATGATCGGTTACGCCACCCCTTGCTCAACGATGCTCTATGGATTAAAGGACCAAGCAG 1096
Qy 541 ATGAAGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATGCTGCTGGGACGCTGTTAGCA 600
Db 1097 ATGAAGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATGCTGCTGGGACGCTGTTAGCA 1156
Qy 601 TCGGCGAGGCAATGGC 615
Db 1157 TCGGCGAGGCAATGGC 1171

RESULT 2
AAA48442
ID AAA48442 standard; DNA; 1231 BP.
XX
AC AAA48442;
XX
DT 08-SEP-2000 (first entry)
XX
DE *E. coli* L-homoserine resistance gene, *rhtB*.

XX L-homoserine resistance; L-homoserine synthesis; *rhtB*; L-threonine;
KW L-valine; L-leucine; ds.
XX
OS *Escherichia coli*.
PH Key Location/Qualifiers
FT 557-1174
FT CDS
FT /tag= a
FT /product= "RhtB"
XX
PN EF1013765-A1.
XX
XX 28-JUN-2000.
XX
XX 20-DEC-1999; 99EP-00125406.
XX
XX 23-DEC-1998; 98RU-00123511.
XX
XX (AJIN) AJINOMOTO KK.
XX
XX Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;
PI
XX WPI; 2000-414802/36.
DR P-PSDB; AAY99597.
XX
XX Novel *Escherichia* bacterium having enhanced L-threonine resistance due to
PT enhanced RhtC protein activity, used to produce L-threonine, L-
PT homoserine, L-valine and L-leucine.
XX
XX Claim 4; Page 12-13; 24pp; English.
XX
XX The present sequence is the L-homoserine resistance gene, *rhtB*, from
CC *Escherichia coli*. This sequence may be used to impart L-homoserine
CC resistance on *E. coli* bacteria, which would be useful for producing a
CC high yield of L-homoserine. L-homoserine resistance means that the
CC bacteria will be able to grow on a minimal medium containing L-homoserine
CC at a concentration at which the corresponding wild-type strain would not
CC grow. Since the transformed bacteria can grow on the minimal medium, it
CC can synthesise L-homoserine, which accumulates. The accumulated amino
CC acids can then be removed from the culture medium. The bacterium of the
CC present invention may also be used to synthesise L-threonine, L-valine
CC and L-leucine at increased levels
XX
SQ Sequence 1231 BP; 291 A; 319 C; 311 G; 310 T; 0 U; 0 Other;
Query Match 100.0%; Score 615; DB 3; Length 1231;
Best Local Similarity 100.0%; Pred. No. 2e-193;
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTAAACGCTGCGCA 60
Db 557 ATGACCTTAGAATGGTGGTTTGCTACCTGCTGACATCGATCATTTAAACGCTGCGCA 616
Qy 61 GGCTCTGGTCAATCAACACTATGACCACTGCTGCTCAACACAGGTATCCGCCGGTGGC 120
Db 617 GGCTCTGGTCAATCAACACTATGACCACTGCTGCTCAACACAGGTATCCGCCGGTGGC 676
Qy 121 GTCTATTGCTGGGCTTCAGACCGGACTGCGATTTCATATTGTGCTGGTGGCGTGGGTT 180
Db 677 GTCTATTGCTGGGCTTCAGACCGGACTGCGATTTCATATTGTGCTGGTGGCGTGGGTT 736
Qy 181 GGGACGCTATTTTCCCGCTCAGTTCGTTTGAAGTGTGAAGTGGGACGCGCGCT 240
Db 737 GGGACGCTATTTTCCCGCTCAGTTCGTTTGAAGTGTGAAGTGGGACGCGCGCT 796
Qy 241 TACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATTAATCG 300
Db 797 TACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGAATTAATCG 856
Qy 301 CTGGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 360
Db 857 CTGGGCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTCACC 916

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (I) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in

RESULT 4	
AA889415/C	
ID	AA889415 standard; cDNA; 2216 BP.
XX	
XX	AA889415;
XX	
XX	13-FEB-2002 (first entry)
XX	
XX	DNA encoding novel human diagnostic protein #25219.
DE	
XX	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
XX	

CC by which to identify the causative agents of a bacterial infection. Also,
CC antibodies generated against proteins translated from mRNA transcribed
CC from proliferation-required sequences can also be used to screen for
CC specific microorganisms that produce such proteins in a species-specific
CC manner. AAH84371 and AAH84670 represent sequencing primers used in the
CC isolation of *E. coli* growth and proliferation related sequence, which are
CC used in an example from the present invention

XX Sequence 417 BP; 87 A; 96 C; 116 G; 118 T; 0 U; 0 Other;

Query Match 67.3%; Score 414; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 7.6e-127;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 GTGATTCGGTTGAAGTGTGAAGTGGGAGGCGGCTTACTTGAATTTGGCTGGGAATC 261
DB 1 GTGATTCGGTTGAAGTGTGAAGTGGGAGGCGGCTTACTTGAATTTGGCTGGGAATC 60
QY 262 CAGCAGTGGCGCGCGCTGGTGCATTTGATTCACCTTAATCGCTGGCTTACTCAATCGCGT 321
DB 61 CAGCAGTGGCGCGCGCTGGTGCATTTGATTCACCTTAATCGCTGGCTTACTCAATCGCGT 120
QY 322 CGACATTTGTTCCAGGCGCGAGTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTT 381
DB 121 CGACATTTGTTCCAGGCGCGAGTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTT 180
QY 382 CTGGCGGCGCTATTCCGCAATTCATCATGCGCGAAGCGCGCAACTGATGATATATC 441
DB 181 CTGGCGGCGCTATTCCGCAATTCATCATGCGCGAAGCGCGCAACTGATGATATATC 240
QY 442 GTGCTGGCGCTACCACTATTGGTGGCGATATTATTGTGATGATCGGTTACGCCACCTT 501
DB 241 GTGCTGGCGCTACCACTATTGGTGGCGATATTATTGTGATGATCGGTTACGCCACCTT 300
QY 502 GCTCAACGAGTTCCTATGAGTTAAAGGACCAAGCGAGATGAAGGCGCTGAATAAGATT 561
DB 301 GCTCAACGAGTTCCTATGAGTTAAAGGACCAAGCGAGATGAAGGCGCTGAATAAGATT 360
QY 562 TTCGGCTCTGTTTATGCTGTGGGAGCGCTGTAGCATGGCGAGGCGATCG 615
DB 361 TTCGGCTCTGTTTATGCTGTGGGAGCGCTGTAGCATGGCGAGGCGATCG 414

RESULT 6

ACA18626

ID ACA18626 standard; DNA; 417 BP.

XX ACA18626;

AC ACA18626;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #283.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Escherichia coli.

XX WO200271183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (BLIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX

PI

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
DR P-PSDB; ABU14756.

XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 6496; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 417 BP; 87 A; 96 C; 116 G; 118 T; 0 U; 0 Other;

Query Match 67.3%; Score 414; DB 7; Length 417;

Best Local Similarity 100.0%; Pred. No. 7.6e-127;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 GTGATTCGGTTGAAGTGTGAAGTGGGAGGCGGCTTACTTGAATTTGGCTGGGAATC 261
DB 1 GTGATTCGGTTGAAGTGTGAAGTGGGAGGCGGCTTACTTGAATTTGGCTGGGAATC 60
QY 262 CAGCAGTGGCGCGCGCTGGTGCATTTGATTCACCTTAATCGCTGGCTTACTCAATCGCGT 321
DB 61 CAGCAGTGGCGCGCGCTGGTGCATTTGATTCACCTTAATCGCTGGCTTACTCAATCGCGT 120
QY 322 CGACATTTGTTCCAGGCGCGAGTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTT 381
DB 121 CGACATTTGTTCCAGGCGCGAGTTTGTGAATCTCACCAATCCCAAAAGTATTGTGTTT 180
QY 382 CTGGCGGCGCTATTTCGCAATTCATCGCGCAACAGCGCGCAACTGATGATATATC 441
DB 181 CTGGCGGCGCTATTTCGCAATTCATCGCGCAACAGCGCGCAACTGATGATATATC 240
QY 442 GTGCTGGCGGTCACCACTATTGTGCTGATATATTGTGATGATCGGTTACGCCACCTT 501
DB 241 GTGCTGGCGGTCACCACTATTGTGCTGATATATTGTGATGATCGGTTACGCCACCTT 300
QY 502 GCTCAACGAGTTCCTATGAGTTAAAGGACCAAGCGAGATGAAGGCGCTGAATAAGATT 561
DB 301 GCTCAACGAGTTCCTATGAGTTAAAGGACCAAGCGAGATGAAGGCGCTGAATAAGATT 360

PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU32233.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids, required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 23973; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Query Match 61.0%; Score 375; DB 7; Length 618;
Best Local Similarity 77.6%; Pred. No. 8.6e-114;
Matches 479; Conservative 0; Mismatches 135; Indels 3; Gaps 2;
1 ATGACCTTATGATGGTGGTTCCCTACTCTGCTGACATCATATTTAAACCTCTGCCCA 60
1 ATGACCATTTGAGTGGTGGTTCCCTACTCTGCTGACATCATATTTAAACCTCTGCCCG 60
61 GGCTCTGGTGGCAATCAACTATGACCACTCGCTCAACACGGTTA--TCGGCCGGTG 118
61 GGTTCCGGAGCGATTTATACCTATGACCTTCACTTACACGATACCGGGCGGGCG 120
119 GCCTCTATTCCTGGGCTTCAGACCGGATGGCGATTCATATTCGCTGGTGGCG-TGGG 177
121 GCCTCGATTCGGGCTTGACAGACCGGGCTGGTTATCATATCGTCTGGTGGCGTTGGT 180

QY 178 GTTGGGACGCTATTTTCCCGCTCAGTGAATTCGCTTTGAAGTGTGAAGTGGGACGGCG 237
DB 181 CTCGGCACTCTCTCTCCCGCTCGGTGCTGGCTTTGAGGTGCTGAAATGGCGCGCGCC 240
QY 238 GCTTACTTGTATTTGGCTGGGATCCACAGCTGGGCGCCGCTGTGCAATTGACCTTAA 297
DB 241 GCCTACCTGATCTGGCTGGGATCCAGCAATGGCGGCGGACGAGGCGATCGACCTGAAT 300
QY 298 TCGTGGGCTCTACTCAATCGCGTCCGACATTTGTTCCAGCGCGCAGTTTTTGTGAATCTC 357
DB 301 ACGTGGCGGAGGCGCAGACGCGCGGCAAACTGTTCCAGCGCGGTGTTGTAATCTC 360
QY 358 ACCAATCCAAAAGTATGTGTTTCGGCGGCGCTATTTCCGCAATTCATCATCGCGCAA 417
DB 361 ACCAATCCAAAAGCATTTGTTTTCGGCGCGCTGTTTCCGAGTTTATCTCGCGCGAC 420
QY 418 CAGCGCACTGATGTCAGTATATCGTCTCGGGGTACACACTATTTGTGGTCGATATATT 477
DB 421 CAGCGCGAGTGTGATGATATCTGTTCTGGGGTACACCACTCGTCTCGTATCAATT 480
QY 478 GTGATGATCGGTTACGCCACCCCTTGTCTCAACGAGTTCCTCTATGATTAAGGACCAAG 537
DB 481 GTGATGATGCTGATGCGACCCCTCGCGCAGCGCATTTCCGCTGATCAAAAGGCGCAAAG 540
QY 598 GCATCGCGGAGGCGCATGC 614
DB 601 GCCTCGCGCGCGCATGC 617
RESULT 9
ACA32157
ID ACA32157 standard; DNA; 618 BP.
XX ACA32157;
XX 19-JUN-2003 (first entry)
XX DE Prokaryotic essential gene #13814.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX KW drug design; gene.
XX OS Enterobacter cloacae.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX P-PSDB; ABU28287.
XX DR New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX

PS Claim 14; SEQ ID NO 20027; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

CC prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

SQ Sequence 618 BP; 106 A; 176 C; 183 G; 153 T; 0 U; 0 Other;

Query Match 58.1%; Score 357.4; DB 7; Length 618;

Best Local Similarity 75.9%; Pred. No. 6.1e-108;

Matches 468; Conservative 0; Mismatches 146; Indels 3; Gaps 2;

QY 1 ATGACCTTGAATGGTGGTTTCCTACCTGCTGACATCGATCATTTTAAAGCTGCGCCA 60

DB 1 ATGACCTTGAATGGTGGTTTCCTACCTGCTGACATCGATCATTTTAAAGCTGCGCCA 60

QY 61 GGCTCTGGTGCATCAACACTATGACACCTCGCTCAACACCGGTTA--TCGGCCGGGT 118

DB 61 GGCTCTGGTGCATTAACACCTATGACACCTCGCTCAACACCGGTTA--TCGGCCGGGT 120

QY 119 GCCTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGCTGGTGGGT--GGG 177

DB 121 GCATCGATTGCCGGTTTGACAGACCGGCTGGGATTTCATATTGCTGGTGGGTATCGGT 180

QY 178 GTTGGACGCTATTTTCCCGCTCAGTGAATGCTGGTTGAAGTGTGAAGTGGCAGGCGCG 237

DB 181 CTGGGACGCTGTCTCCCGCTCCGCTGGCTTGAAGTGTGAAGTGGCAGGCGCGCA 240

QY 238 GCTTACTTCTGATTGGTGGGAATCCAGAGTGGCGCGCTGGTGAATGACCTTAAA 297

DB 241 GCGTATCTGATTGGTGGGATCCAGAGTGGCGCGCGGAGGCTCCATCAACTGAAT 300

QY 298 TGCTGGCTCTTACTCAATTCGCTGCATTTGTTCCAGCGCGCAGTTTGTGAATCTC 357

DB 301 AGCTGGCTCTGAGCAAAACCGCGCCATCTGTTTAAAGCGTGGGATTTCGTCACCTG 360

QY 358 ACCAATCCCAAGATTTGTTTTCGGCGGCTATTTCGGCAATTCATCATCGCGCA 417

DB 361 ACCAATCCCAAGATTTGTTTTCGGCGGCTATTTCGGCAATTCATCATCGCGCA 420

QY 418 CAGCCCAACTGATGAGTATATCGTGGTCCGGCGTCAACACTATTGGTGCATATTATT 477

DB 421 CAGCCCTCAGTGATGAGTACGTGGTGGTGGCGGACCACTATTATGCTGATCATC 480

QY 478 GTGATGATCGTTCACGCCACCTTGTCTCAACGATTTGCTCTATGATTAAGGACCAAG 537

DB 481 GTGATGATTGGTTACGCGACGCTGGCGCAGCAATTTCGCGGTGAATTAAGGCGCTAAG 540

QY 538 CAGATGAAGCGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATGCTGGGAGCGCTGTTA 597

DB 541 CAGATGAAGCGCGCTGAATAAGATTTTCGGCTCGTGTGTTTATGCTGGGAGCGCTGCT 600

QY 598 GCATCGCGCGGCGATGC 614

DB 601 CGCTCAGCGCGTCACGC 617

RESULT 10

ACH48809

ID ACA48809 standard; DNA; 623 BP.

XX ACA48809;

AC ACA48809;

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #30466.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; gene.

XX Salmonella paratyphi.

OS WO200277183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362999P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU44939.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 36579; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 623 BP; 124 A; 171 C; 168 G; 160 T; 0 U; 0 Other;
Query Match 56.1%; Score 344.8; DB 7; Length 623;
Best Local Similarity 76.6%; Pred. No. 9.5e-104;
Matches 475; Conservative 0; Mismatches 137; Indels 8; Gaps 4;
QY 1 ATGACCTTAGAATGGTGGTGGCTTACCTGCTGATCGATCATTTTAACGCTGTCGCCA 60
Db 1 ATGACCTTAGAATGGTGGTGGCTTACCTGCTGATCGATCATTTTAACGCTGTCGCC 60
QY 61 GCGCTGTGGTGAATCAACACTATGACCACTCGCTCAACCAACGCTTATCCGGCGC--GTG 118
Db 61 GGTTCAGCGCGCATCAATACCATGACGAGCTCTATCAACCATGATATCGTGGCGAGCG 120
QY 119 GCGTCTATTGCTGGCTTCAGACCGGACTGCGGATTCATATTGCTGGTTCGCTGGCG 178
Db 121 GCTTCTATCGCCGACTCCACCGGACTGGGATACATATCGTACTGCTGGCGCTCGGA 180
QY 179 TTGGG-AGCTATTATTCCTCGCTGATGATTCGTTGAAGTGTGAGTGGCGAGCGCG 237
Db 181 CTGGGTAGCTCTTTTCGCTGCTGCTCTGCTTTTGAATTCGAAATGGCTGGCGG 240
QY 238 GCTTACTTGAATTTGCTGGGATTCAGAGTGGCGCGCGTGGTGAATGACCTTAA 297
Db 241 GCTTATCTATCTGCTGGGTATCCGCAATGGCG---CGACGCGCTATCGATTCGCAT 297
QY 298 TCGCTGGGCTCTACTCAATCGCTCGCATTTGTTCCAGCGCGCAGTTTGTGGAATCTC 357
Db 298 ACTCTCGCCAGACGCAATCGCGGCTGCTGTTTCAACGCGGATTTTGTCAATCTA 357
QY 358 ACCAATCCCAAAAGATTGTTGTTTTCGGCGCGCTATTTCGCAATTCATCGCGCAA 417
Db 358 ACCAATCCCAAAAGATTGTTGTTTTCGGCGCGCTATTTCGCAATTCATCGCGCGAG 417
QY 418 CAGCGCAACTGATGCAATGATGCTGCTGGGCTCACCATTGTTGGTTCGATATTAT 477
Db 418 CAAACGCACTGGCGAGTACTCATTTCTGCGGTACACAGTTGGTGGATATGATT 477
QY 478 GTGATGATCGTTAGCCACCCCTTGCCTCAACGG--ATTGCTCTATGATTAAGGACCAA 535
Db 478 GTGATGACCGGTTAGCCACACTGGCGTAGCGCACTTTCGCGCTGATTAAGGACCAA 537
QY 536 AGCAGATGAAGCGCTGAATGAATTTTCGGCTGTTGTTTATGCTGGTGGAGCGCTGT 595
Db 538 AGCAGATGAAGCGCTGAATGAATGCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 597
QY 596 TAGCATCGCGGAGGATCGG 615
Db 598 TGGCGTGGCAAGACACGCG 617

RESULT 11
ACA53970
ID ACA53970 standard; DNA; 621 BP.
XX
AC ACA53970;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #35627.
XX

KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Yersinia pestis.
XX
PN WO200277183-A2.
XX
XX 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR P-PSDB; ABU50100.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 41840; 1766pp; English.
CC
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 621 BP; 135 A; 143 C; 156 G; 187 T; 0 U; 0 Other;
Query Match 44.7%; Score 274.8; DB 7; Length 621;
Best Local Similarity 67.5%; Pred. No. 1.8e-80;
Matches 416; Conservative 0; Mismatches 197; Indels 3; Gaps 2;
QY 1 ATGACCTTAGAATGGTGGTGGCTTACCTGCTGATCGATCATTTTAACGCTGTCGCCA 60
Db 1 ATGACCTTAGAATGGTGGTGGCTTACCTGCTGATCGATCATTTTAACGCTGTCGCC 60

AC AAS89412;
XX
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #25216.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
PN W0200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSRQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG25225.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 25216; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1884 BP; 495 A; 464 C; 479 G; 446 T; 0 U; 0 Other;
SQ
Query Match 40.4%; Score 248.6; DB 5; Length 1884;
Best Local Similarity 96.6%; Pred. No. 1.6e-71;
Matches 254; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 320 GTCGACATTTGTTCCAGCGCGAGTTTGTGAATCTCACCACATCCCAAAAGTATTGTGT 379
DB 1160 GTCCGACTATGATATATTCGGCAGTTTGTGAACTCACCACATCCCAAAAGTATTGTGT 1219
QY 380 TTCTGGCGCGCTATTTCCGCAATTCATATCCGCAACAGCCGCAATGATGAGTATA 439
DB 1220 TTCTGGCGCGCTATTTCCGCAATTCATATCCGCAACAGCCGCAATGATGAGTATA 1279
QY 440 TCGTGTCTCGCGCTACCATTTTGTGTCGATTTATTTGTCATCGGTTACGCCACCC 499
DB 1280 TCGTGTCTCGCGCTACCATTTTGTGTCGATTTATTTGTCATCGGTTACGCCACCC 1339

QY 500 TTGCTCAACGGATTGCTCTATGATTAAAGACCAAAAGCAGATGAAGCGCTGAATAAGA 559
DB 1340 TTGCTCAACGGATTGCTCTATGATTAAAGACCAAAAGCAGATGAAGCGCTGAATAAGA 1399
QY 560 TTTTCGGCTCGTTGTTTATGCTG 582
DB 1400 TTTTCGGCTCGTTGTTTATGCTG 1422

RESULT 14
AAH84479/c
ID AAH84479 standard; DNA; 263 BP.
XX
XX AAH84479;
AC
XX
DT 26-SEP-2001 (first entry)
XX
DE E. coli growth and proliferation related DNA sequence SEQ ID NO:107.
XX
XX Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW bacterial infection; microorganism; ds.
XX
XX Escherichia coli.
OS
XX W0200134810-A2.
PN
XX
PD 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US030950.
PF
XX 09-NOV-1999; 99US-0164415P.
PR
XX (ELIT-) ELITRA PHARM INC.
PA
XX Forsyth RA, Ohlsen K, Zyekind J;
PI
XX WPI; 2001-335933/35.
DR
XX
XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT for screening for homologous genes and for designing expression vectors.
XX
XX Claim 1; Page 126; 522pp; English.
PS
XX
XX AAH44373 to AAH4499 represent Escherichia coli growth and proliferation
CC related DNA sequences (I). AAH4500 to AAH4670 encode the E. coli growth
CC and proliferation related proteins given in AAG99078 and AAG99830 to
CC AAG98999. (I) can be used as potential targets for the generation of new
CC antimicrobial agents, and for identification of compounds which interact
CC with the gene products of (I). In addition the expression of (I) and the
CC purification of the proteins, the purified proteins can be used to
CC generate reagents and screen small molecule libraries or other candidate
CC compound libraries for compounds that can be further developed to yield
CC novel antimicrobial compounds. In addition, nucleic acid probes
CC complementary to (I) that are specific for particular species of
CC microorganisms can be used to identify particular microorganism species
CC in clinical specimens, therefore, providing a rapid and dependable method
CC by which to identify the causative agents of a bacterial infection. Also,
CC antibodies generated against proteins translated from mRNA transcribed
CC from proliferation-related sequences can also be used to screen for
CC specific microorganisms that produce such proteins in a species-specific
CC manner. AAH4371 and AAH4670 represent sequencing primers used in the
CC isolation of E. coli growth and proliferation related sequence, which are
CC used in an example from the present invention
XX
XX Sequence 263 BP; 72 A; 60 C; 68 G; 63 T; 0 U; 0 Other;
SQ
Query Match 33.7%; Score 207.4; DB 4; Length 263;
Best Local Similarity 94.3%; Pred. No. 2.9e-58;
Matches 248; Conservative 0; Mismatches 11; Indels 4; Gaps 3;
QY 305 CCTCTACTCAATCGGTCGACATTTGTTCCAGCGCGAGTTTGTCAATTCACCAATC 364

Db 263 CCCCTACTCAATCGCTGACATTTGTTCCAGCGCGCAGTTTTTTGTGTAATCTCACCATC 204
Qy 365 CCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATGCGCAACAGCGGC 424
Db 203 CCAAAAGTATTGTGTTTCTGGCGGCGCTATTTCCGCAATTCATCATGCGCAACAGCGGC 144
Qy 425 AACTGATGCAATATATCGTCTC-GCGTCCACACTATTTGGTGTGATATTATTTGATG 483
Db 143 AACTGATGCAATATATCGTCTC-GCGTCCACACTATTTGGTGTGATATTATTTGATG 84
Qy 484 ATC-GGTTAGCGCACCCCTTGTCTCAACGGATTG--CTCTATGGATTAAAGGACCAAGCAG 540
Db 83 ATCGGGTTACGCCACCCCTTGTCTCAACGGATTGCTCTAAGGGATTAAAGGCCCAACCA 24
Qy 541 ATGAGGCGCTGAATAGATTTT 563
Db 23 ATGAGGCGCTGAATAGATTTT 1

RESULT 15

AAS89410/c

ID AAS89410 standard; cDNA; 274 BP.

XX

AC AAS89410;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #25214.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN W0200175067-A2.

XX

PD 11-OCT-2001.

XX

XX 30-MAR-2001; 2001WO-US008631.

XX

XX 31-MAR-2000; 2000US-00540217.

XX

XX 23-AUG-2000; 2000US-00649167.

XX

PA (HYSE-) HYSEQ INC.

XX

XX Drmanac RT, Liu C, Tang YT;

XX

XX WPI; 2001-639362/73.

XX

XX P-PSDB; ABG25223.

XX

XX New isolated polynucleotide and encoded polypeptides, useful in

XX

XX diagnostics, forensics, gene mapping, identification of mutations

XX

XX responsible for genetic disorders or other traits and to assess

XX

XX biodiversity.

XX

XX Claim 1; SEQ ID NO 25214; 103pp; English.

XX

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

XX

XX sequences. (I) is useful as hybridisation probes, polymerase chain

XX

XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

XX

XX and in recombinant production of (II). The polynucleotides are also used

XX

XX in diagnostics as expressed sequence tags for identifying expressed

XX

XX genes. (I) is useful in gene therapy techniques to restore normal

XX

XX activity of (II) or to treat disease states involving (II). (II) is

XX

XX useful for generating antibodies against it, detecting or quantitating a

XX

XX polypeptide in tissue, as molecular weight markers and as a food

XX

XX supplement. (II) and its binding partners are useful in medical imaging

XX

XX of sites expressing (II). (I) and (II) are useful for treating disorders

XX

XX involving aberrant protein expression or biological activity. The

XX

XX polypeptide and polynucleotide sequences have applications in

XX

XX diagnostics, forensics, gene mapping, identification of mutations

XX

XX responsible for genetic disorders or other traits to assess biodiversity

XX

XX and to produce other types of data and products dependent on DNA and

XX

CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: the sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 274 BP; 61 A; 71 C; 70 G; 72 T; 0 U; 0 Other;

Query Match 28.5%; Score 175; DB 5; Length 274;

Best Local Similarity 100.0%; Pred. No. 1.7e-47;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 409 ATGCCGCAACAGCCGCAACTGATGCAATATATCGTCTCGSCGTCACCACTATTGTGTC 468

Db 274 ATGCCGCAACAGCCGCAACTGATGCAATATATCGTCTCGSCGTCACCACTATTGTGTC 215

Qy 469 GATATTATTGTGATGATCGTTACGCCACCTTCTCTCAACGGATTGCTCTATGGAATAA 528

Db 214 GATATTATTGTGATGATCGTTACGCCACCTTCTCTCAACGGATTGCTCTATGGAATAA 155

Qy 529 GGACCAAGACAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGG 583

Db 154 GGACCAAGACAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGG 100

Search completed: March 28, 2004, 14:38:03

Job time : 310.314 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:10:49 ; Search time 61.6694 Seconds
(without alignments)
5534.263 Million cell updates/sec

Title: US-09-847-392-1_COPY_557_1171

Perfect score: 615
Sequence: 1 atgaccttagatggtggtt.....tagatcgcgagcatgog 615

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
 - 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
 - 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
 - 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
 - 5: /cgn2_6/ptodata/2/ina/PCURS_COMB.seq:*
 - 6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	615	100.0	1200	US-09-396-357-1	Sequence 1, Appli
2	444	67.3	417	US-09-711-184-274	Sequence 274, App
3	375	61.0	645	US-09-489-039A-895	Sequence 895, App
C 4	207.4	33.7	263	US-09-711-164-107	Sequence 107, App
5	112.4	18.3	834	US-09-252-991A-6447	Sequence 6447, Ap
C 6	102.4	16.7	930	US-09-252-991A-6103	Sequence 6103, Ap
7	75.4	12.3	653	US-09-328-352-777	Sequence 777, App
8	59	9.6	645	US-09-489-039A-2622	Sequence 2622, Ap
9	37.2	6.0	633	US-09-328-352-2682	Sequence 2682, Ap
10	36	5.9	312	US-09-543-681A-3779	Sequence 3779, Ap
11	35.4	5.8	1028	US-09-655-908-17	Sequence 17, Appl
12	32.8	5.3	1442	US-09-634-238-82	Sequence 82, Appl
13	32.2	5.2	681	US-09-252-991A-15039	Sequence 15039, A
C 14	32	5.2	1001	US-09-641-638-278	Sequence 278, App
C 15	32	5.2	1597	US-09-634-238-195	Sequence 195, App
C 16	31.8	5.2	738	US-09-252-991A-6015	Sequence 6015, Ap
17	31.8	5.2	1128	US-09-252-991A-6376	Sequence 6376, Ap
18	31.8	5.2	2136	US-09-252-991A-6294	Sequence 6294, Ap
C 19	31.8	5.2	3948	US-09-252-991A-6175	Sequence 6175, Ap
20	31.2	5.1	505	US-09-621-976-15639	Sequence 15639, A
21	31	5.0	540	US-09-489-039A-5266	Sequence 5266, Ap
22	31	5.0	726	US-09-489-039A-2390	Sequence 2390, Ap
23	31	5.0	1179	US-09-489-039A-2656	Sequence 2656, Ap
C 24	30.6	5.0	351	US-09-543-681A-2733	Sequence 2733, Ap
C 25	30.6	5.0	1137	US-09-252-991A-6183	Sequence 6183, Ap
26	30.6	5.0	1698	US-09-489-039A-5833	Sequence 5833, Ap
C 27	30.4	4.9	1014	US-09-325-932A-15	Sequence 15, Appl

C 28	30.4	4.9	1560	4	US-09-252-991A-11778	Sequence 11778, A
29	30.4	4.9	2157	4	US-09-252-991A-11658	Sequence 11658, A
30	30.4	4.9	2187	4	US-09-252-991A-11598	Sequence 11598, A
C 31	30.4	4.9	2460	4	US-09-328-352-1216	Sequence 1216, Ap
32	30.2	4.9	636	4	US-09-252-991A-4176	Sequence 4176, Ap
C 33	30.2	4.9	1071	4	US-09-252-991A-4432	Sequence 4432, Ap
34	30.2	4.9	1623	4	US-09-513-783A-33	Sequence 33, Appl
35	30.2	4.9	2553	4	US-09-252-991A-4320	Sequence 4320, Ap
36	30	4.9	440	4	US-09-621-976-12898	Sequence 12898, A
C 37	29.8	4.8	1128	4	US-09-161-994A-1	Sequence 1, Appli
38	29.8	4.8	1173	4	US-09-543-681A-1224	Sequence 1224, Ap
39	29.8	4.8	7218	1	US-08-232-463-14	Sequence 14, Appl
C 40	29.8	4.8	4403785	3	US-09-103-840A-2	Sequence 2, Appli
C 41	29.8	4.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
42	29.6	4.8	438	4	US-09-489-039A-595	Sequence 595, App
C 43	29.6	4.8	2623	2	US-08-973-675-1	Sequence 1, Appli
C 44	29.6	4.8	3925	2	US-09-047-026A-3	Sequence 3, Appli
C 45	29.6	4.8	4859	4	US-09-453-702B-154	Sequence 154, App

ALIGNMENTS

RESULT 1
US-09-396-357-1
; Sequence 1, Application US/09396357
; Patent No. 6303348
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMIOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHAKOVA, IRINA LJOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/396,357
; CURRENT FILING DATE: 1999-09-15
; EARLIER APPLICATION NUMBER: RU98118425
; EARLIER FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (557)..(1171)
US-09-396-357-1

Query Match	100.0%	Score 615;	DB 4;	Length 1200;
Best Local Similarity	100.0%	Pred. No. 7.3e-200;	Mismatches 0;	Indels 0;
Matches 615;	Conservative 0;			Gaps 0;
QY	1	ATGACCTTAGAATCGTGGTTGGCTACCTCTGACATCGATCAATTTAAAGCTGTGCGCA	60	
Db	557	ATGACCTTAGAATCGTGGTTGGCTACCTCTGACATCGATCAATTTAAAGCTGTGCGCA	616	
QY	61	GGCTCTGCTGAATCAACATCATGACCACTCGCTCAACCGGTTATCCGGCCGGTGGC	120	
Db	617	GGCTCTGCTGAATCAACATCATGACCACTCGCTCAACCGGTTATCCGGCCGGTGGC	676	
QY	121	GTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTCTGCTGGCTGGGGTT	180	
Db	677	GTCTATTGCTGGGCTTCAGACCGGACTGGCGATTTCATATTGTCTGCTGGCTGGGGTT	736	
QY	181	GGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGCGGCGGCT	240	
Db	737	GGGACGCTATTTTCCCGCTCAGTGATTGCGTTTGAAGTGTGAAGTGGGCGGCGGCT	796	
QY	241	TACTTGTATTGGCTGGGAATCCAGACGTGGCGGCGGCTGGTGCATTTGACCTTAAATCG	300	

Db 797 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCAATTGACCTTAAATCG 856
Qy 301 CTGGCCTCTACTCAATCCGCTGCAGCATTTGTTCCAGCGCGAGTTTGTGAATCTCACC 360
Db 857 CTGGCCTCTACTCAATCCGCTGCAGCATTTGTTCCAGCGCGAGTTTGTGAATCTCACC 916
Qy 361 AATCCCAAAAGTAGTTGTTTCTGGCGCGCTTAATTTCCGCAATTCATCATCCGCAACAG 420
Db 917 AATCCCAAAAGTAGTTGTTTCTGGCGCGCTTAATTTCCGCAATTCATCATCCGCAACAG 976
Qy 421 CCGCAACTGATGCAGTATATCGTGCAGCTGCAGCATTTGTTGGTGATATTATTGTG 480
Db 977 CCGCAACTGATGCAGTATATCGTGCAGCTGCAGCATTTGTTGGTGATATTATTGTG 1036
Qy 481 ATGATCGGTTACGCCACCTTGTCTCAACGAGTGTCTCTATGGAATTAAGAGCAACAGCAG 540
Db 1037 ATGATCGGTTACGCCACCTTGTCTCAACGAGTGTCTCTATGGAATTAAGAGCAACAGCAG 1096
Qy 541 ATGAAGCGCTGAATGAAGATTTTCGCTCGTTGTTTATGCTGGGAGCGCTGTAGCA 600
Db 1097 ATGAAGCGCTGAATGAAGATTTTCGCTCGTTGTTTATGCTGGGAGCGCTGTAGCA 1156
Qy 601 TCGCGAGGAGCATCGG 615
Db 1157 TCGCGAGGAGCATCGG 1171

RESULT 2
US-09-711-164-274
; Sequence 274, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Chisen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELIPIRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 274
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(417)
US-09-711-164-274

Query Match 67.3%; Score 414; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.9e-131; Indels 0; Gaps 0;
Matches 414; Conservative 0; Mismatches 0;

Qy 202 GTGATTCGTTGAAGTGTGAAGTGGCAGCGCGCTTACTTGATTGGCTGGGAATC 261
Db 1 GTGATTCGTTGAAGTGTGAAGTGGCAGCGCGCTTACTTGATTGGCTGGGAATC 60
Qy 262 CAGCAGTGGCGCGCTGTGTGAATGACCTTAATTCGCTGGCCTCTACTCAATCGGCT 321
Db 61 CAGCAGTGGCGCGCTGTGTGAATGACCTTAATTCGCTGGCCTCTACTCAATCGGCT 120
Qy 322 CGACATTTGTTCCAGCGCGCAGTTTGTGAAATCTCACCATCCCAAAAGTATTGTGTT 381
Db 121 CGACATTTGTTCCAGCGCGCAGTTTGTGAAATCTCACCATCCCAAAAGTATTGTGTT 180
Qy 382 CTGGCGCGCTATTTCGGCAATTCATGTCGCGCAACAGCCGCACTGATGAGTATATC 441
Db 181 CTGGCGCGCTATTTCGGCAATTCATGTCGCGCAACAGCCGCACTGATGAGTATATC 240

Qy 442 GTGCTCGCGCTCACCACTATTGTGTCGATATATTGTGATCGTTACGCCACCCCTT 501
Db 241 GTGCTCGCGCTCACCACTATTGTGTCGATATATTGTGATCGTTACGCCACCCCTT 300
Qy 502 GCTCAACGAGTGTCTATGCAATTAAGGACCAAGCAGATGAGCGCTGAATAAGATT 561
Db 301 GCTCAACGAGTGTCTATGCAATTAAGGACCAAGCAGATGAGCGCTGAATAAGATT 360
Qy 562 TTGCGCTCGTTGTTTATGCTGCTGGAGCGCTGTAGCATCGGCGAGGCATGCG 615
Db 361 TTGCGCTCGTTGTTTATGCTGCTGGAGCGCTGTAGCATCGGCGAGGCATGCG 414

RESULT 3
US-09-489-039A-895
; Sequence 895, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 895
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-895

Query Match 61.0%; Score 375; DB 4; Length 645;
Best Local Similarity 77.6%; Pred. No. 5.1e-118; Indels 3; Gaps 2;
Matches 479; Conservative 0; Mismatches 135;

Qy 1 ATGACCTTAGAATGGTGTGCTTACCTGCTGACATCGATCATTTTAAAGCTGTGCCCA 60
Db 25 ATGACCAATGAGTGTGTGCTGCTGCTGACATCAATTTCTCAGCTGTGCCG 84
Qy 61 GGCTCTGGTGCAATCAACACTATGACCACTCGCTCAACCAAGCTTA--TCCGGCCGGTG 118
Db 85 GGTTCGGAGCGAATTAATACCATGACCACTCCATTAAACCAAGATACCGCGCGCGCG 144
Qy 119 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGGCTCATATTGTGCTGGTGGG-TGGG 177
Db 145 GCGTCGATTCGGGTTTCGACACTGGGCTGGCCATTCTATCGTGTGTTGGCTTGGT 204
Qy 178 GTTGGAGCGCTATTTTCCGCTCAGTGAATTCGCTTTGAAAGTGTGAAGTGGCGAGCGCG 237
Db 205 CTCGCACTCTCTCTCCGCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 264
Qy 238 GCTTACTTGTGTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCATTTGACCTTAAA 297
Db 265 GCTTACTTGTGTTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCATTTGACCTTAAA 324
Qy 298 TCGTGGGCTCTACTCAATTCGCTCGAATTTGTTCCAGCGCGAGTTTGTGAAATCTC 357
Db 325 ACGCTGGGAGGCGCAGACGCGCGGCAAACTGTTTTCAGCGCGCGCTGCTGCTGCTGCTGCT 384
Qy 358 ACCAATCCCAAAAGTATTGTTTCTGGCGGCTATTTCGCAATTCATCATCGCCGAA 417
Db 385 ACCAATCCCAAAAGTATTGTTTCTGGCGGCTGTTTCTGGCGGCTGTTTCTGGCGGCTGTTT 444
Qy 418 CAGCGCAACTGATCAGTATATGCTGCTGGCGGCTACCACTATTGTTGGTGGATATTATT 477
Db 445 CAGCGCAGTGTGATCAATATATGCTTCTGGGCGTCAACCACTCGTCTGATATCATTT 504
Qy 478 GTGATGATCGGTAGCCACCTTGTCTCAAGGATTTGCTCTATGATTAAGGACCAAG 537
Db 505 GTGATGATTTGGCTATGCGACCTCTGGCAGCGCATTTCCGCTATGATCAAGGCGCAAG 564

; ORGANISM: Pseudomonas aeruginosa
US-09-352-991A-6103

Query Match 16.7%; Score 102.4; DB 4; Length 930;
Best Local Similarity 55.6%; Pred. No. 8.8e-25;
Matches 278; Conservative 0; Mismatches 216; Indels 6; Gaps 4;

QY 16 TGGTTTCCTACCTGCTGACATCGATCATTTTAAAGCTGTCCAGAGGCTCTGGTGCAATC 75
Db 504 TGGTTGCGCTTCTTCCTGCGCTGCTGGGCCATCAGCCTGTCCCGGGGCGCGAGCCATC 445
QY 76 AACACATATGACCACTCGCTCAACCAAGGTATCCGGCCGGTGGC--GTCATTGCTGGG 133
Db 444 GCCTCGATGTCCTCGGGTTCAGTACGGCTTTCGGCAGGCTACTGNAACGGCTGGC 385
QY 134 CTTACAGACCGGATCGGGATTCATATGCTGCTGGTGGCG-TGGGGTTGGAGCGCTATTT 192
Db 384 CTGCAGATCGGCTGGCCCTCGCAGATCGCCATTTGTCCGCGCGGGTGGTGGCTG 325
QY 193 TCCCGCTCAGTATGCTGTTTGAAGTGTGAAGTGGCAGGCGCGCTTACTTTGATTTGG 252
Db 324 CGNACCTCGCATGCTTTTCAGCTGATCAAGTGTTCGGCTGGCTACTCTGGTGTAC 265
QY 253 CTGGGAATCCAGAGTGGC--GGCGCGCTGGTGCATTAACCTTAAATCGCTGGCCTCTA 310
Db 264 CTGGCGTGGCGCAGTGGCAGGCGCGCCACAGCGCTTGAGCACCGATGGCGAACGGCT 205
QY 311 CTCATACG-CGTGCATTTGTTCCAGCGCGCAGTTTTGTGAATCTACCAATCCCAAA 369
Db 204 CTGGGGGACCGTTGACCTTGTGCTGGCTGGTTCCTGGTCAAGCCAGCAATCCCAAG 145
QY 370 AGTATTGTTTCTGGCGGCGCTATTTCGCAATTCATCATCGCAGCAACAGCGCACTG 429
Db 144 GCGGTGATCTTCATGCTCGCGGTGCTGCGCAGTTTCATCGACCCGACAGCGCTGCTG 85
QY 430 ATGCAGTATATGCTGCTCGCGTCAACACTATGTGTCGATATTAATTGTGATGTCGT 489
Db 84 GCGCAATACCTGATCATGCGCGCACCATGATGCTGCTGACCTGATGCTATGCGCGGC 25
QY 490 TAGCCACCTTCTCTCAACG 509
Db 24 TACACCGGCTGCTGCGCG 5

RESULT 7
US-09-328-352-777
; Sequence 777, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 777
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-777

Query Match 12.3%; Score 75.4; DB 4; Length 663;
Best Local Similarity 49.1%; Pred. No. 1.2e-15;
Matches 286; Conservative 0; Mismatches 291; Indels 6; Gaps 3;

QY 1 ATGACCTTAGAATGGTGGTTGCTACCTGCTGACATCGATCATTTTAAACGCTGTGCGCA 60
Db 37 ATGTCTTTGCAAGTTGGTTGCTTATATGTTGGCCTGTGGGTCAATTAGTATTCTCCA 96
QY 61 GGCTCTGCTCAATCAACACTATGACCATGCTCAACCGGTTATCCGGCGGTGGC 120
Db 97 GGAGCAGGGGCAATTCGCTTCTATGTCAGCGGCTTAAATATTGGCTTTAGACATGGCTAC 156

QY 121 --GTCTATTGCTGGCTTCAGACCGGACTGCGGATTCATATTGTGCTGGTGGCGTGGG- 177
Db 157 TGGAAATGCAATTTGGTTTGCAAATCGCCTTGTAAATCAAATCATGATTTAGCGCGCGGT 216
QY 178 GTTGGGACGCTATTTTCCCGCTCAGTGAATTCGGTTTGAAGTGTGAAGTGGGACGGCGG 237
Db 217 GTAGGTGTTCTAATTCGCAACGACACCATTAGCTTTTTCAGCGGTTAAGTGGTTTGGGTA 276
QY 238 GGTACTTGTATTGGCTGGGAATCCAGAGTGGCGCGCGCTGGTGCA--ATTGACCTT 294
Db 277 GCTTATTTATTGATTTTAGCCTATTTTGAGTGGACAGCACCTGTAAAGATATAGAAATT 336
QY 295 AATCGCTGGCTCTACTCAATCGCTGCATTTGTTCAGCGCGCAGTTTTTGTGAAT 354
Db 337 CAACAGCAAAAAGAGATTAATCTGTTCTCTCTGCTGTTCAATGGATTGTAGTCAAT 396
QY 355 CTCACCAATCCAAAAGTATTGTGTTTCTGGCGCGCTATTTCGCCAATTCATCATCGG 414
Db 397 ATCAGTAACCTTAAAGCCATCGTATTTTATTGGCAGTTTTCCTCAGTCTTAGATTTA 456
QY 415 CAACAGCGCACTGATCAGTATATCGTCTCGCGTCAACCTATTCAGTATTTGGTCAAT 474
Db 457 AGCAACCTCAATGATACAATATCTGATTATGCGACCACTATGTTAGATTGATTG 516
QY 475 ATTGTGATGATCGGTTAGCCACCTTCTCAACGGATTGCTCTATGCGATTAAAGGACCA 534
Db 517 ATTGTAATGGCTGGTTATACAGGCTAGCTTCAAAGTTTAAAGTTTCTAGCTTCCT 576
QY 535 AAGCAGATGAAGCGCTGAATAAGATTTCGGCTCGTTGTTA 577
Db 577 AAGCAACAAAATATTTAAACCGTGGTTTTCAGCTCATGTTA 619

RESULT 8
US-09-489-039A-2622
; Sequence 2622, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLBBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2622
; LENGTH: 645
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2622

Query Match 9.6%; Score 59; DB 4; Length 645;
Best Local Similarity 46.1%; Pred. No. 4.8e-10;
Matches 197; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 174 TGGGTTGGGACGCTATTTTCCCGCTCAGTGAATTCGGTTTGAAGTGGGACGG 233
Db 204 TGGTTTAGTTCGCTACTGCGGCTTCTGAGTTTGGCTTATACCTTATTGAAATGTCGG 263
QY 234 CCGCGCTTACTTGAATTTGCTGGGAATCCAGCACTGGCGCGCTGTCGCAATTGACCT 293
Db 264 ACAGATACCTTTCCTGGCTGGGATCCAGTTGCTCTGGGACACGCCAGCAATTTAA 323
QY 294 TAAATCGCTGGCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTTTGGAA 353
Db 324 CACCACACCGCGCGAATCAGATTCCAGAGTAATGTTTCTCGTGGGATGCTGGGAA 383
QY 354 TCTACCAATCCCAAGATTTGTTTCTGGCGCGCTATTTCCGCAATTCATCATGCC 413
Db 384 TGTGCTTAACCCCAAAATTTGCTGCTTCTAGCTTTCGTTTACCTCAGTTTATCCCGC 443


```
QY 414 GCAACAGCGCAACTGATGATATATCGTCTCGGCTCACCCTATTTGGTGGATAT 473
Db 444 CGGGCATTTACCCGCTGAGCTGACCTTTCTCTGCTGCTATCCATGATTTGGTAC 503
QY 474 TATTGTGATGATCGGTACGCCACCTTGTCTCAACGGATTCTCTATGGATTAAAGAAC 533
Db 504 GCTGTGCTGTTAAAGCTGATTAACGCCACTCGCTATGCCGACGATCTCTGAAACCA 563
QY 534 AAGCAGATGAGAGCGCTGAATAAGATTTCGGCTCGTGTATTATGCTGGTGGAGCGCT 593
Db 564 AGCCGTAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 623
QY 594 GTTAGCA 600
Db 624 GCTGGCA 630

RESULT 9
US-09-328-352-2682
; Sequence 2682, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2682
; LENGTH: 633
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2682

Query Match 6.0%; Score 37.2; DB 4; Length 633;
Best Local Similarity 46.2%; Pred. No. 0.013;
Matches 123; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 324 ACATTGTTCCAGCGCAGTTTGTGAATCTCACCACCCCAAAAGTATGTTCT 393
Db 348 AAGCTTTTCTATCAAGGTGTTGCTCTACTTAATCCAAAACAATTGCTTTT 407
QY 384 GCGCGGCTATTTCGCAATTCATCAGCCGCAACAGCCGCACTGATGATATCGT 443
Db 408 TATTGCTCTTTTCCACAGTTTATTGATATTAAGAAAGAAATATTAATCAGTTTGGT 467
QY 444 GCTCGCGCTCACCCTATTTGTTGCTGATATTATTGATGATCGTTACGCCCTTGC 503
Db 468 TTATCTTTTAACTTCTGCTGATCGGATTTTAACTTATTGTTTATGCCAATTTTC 527
QY 504 TCAACGGATTGCTCTATGGATTAAAGGACCAAGAGCAGATGAAGCGCTGAATAAGATTTT 563
Db 528 CTCATATTCAAGAGAAGATGCTTGCGGAAATAATTTTCTAAGCTCAATTAAGTCAG 587
QY 564 CGCTCGTTGTTATGCTGGTGGAG 589
Db 588 TGGGTGATATTTTCTTTTAGCAG 613

RESULT 10
US-09-543-681A-3779
; Sequence 3779, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706

QY 414 GCAACAGCGCAACTGATGATATATCGTCTCGGCTCACCCTATTTGGTGGATAT 473
Db 444 CGGGCATTTACCCGCTGAGCTGACCTTTCTCTGCTGCTATCCATGATTTGGTAC 503
QY 474 TATTGTGATGATCGGTACGCCACCTTGTCTCAACGGATTCTCTATGGATTAAAGAAC 533
Db 504 GCTGTGCTGTTAAAGCTGATTAACGCCACTCGCTATGCCGACGATCTCTGAAACCA 563
QY 534 AAGCAGATGAGAGCGCTGAATAAGATTTCGGCTCGTGTATTATGCTGGTGGAGCGCT 593
Db 564 AGCCGTAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 623
QY 594 GTTAGCA 600
Db 624 GCTGGCA 630

Query Match 5.9%; Score 36; DB 4; Length 312;
Best Local Similarity 48.5%; Pred. No. 0.023;
Matches 99; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 361 AATCCAAAAGTATGTGTTTCTGCGCGCTATTTCCGCAATTCATCTCGCGCAACAG 420
Db 52 AATCCAAAAGCAATACTTATCTTTACGGCTTTCTTCCGCAATTCGTTAGCCCAATTA 111
QY 421 CGCAACTGATGATGATATATCGTCTCGGCTCACCCTATTTGGTGGATATTTATGTG 480
Db 112 CGGCTAGTTCCTGATTTCTAATACCTTTGTTCTTTATCTGTTACTCGAGTTTATGCA 171
QY 481 ATGATCGGTTAGCCACCTTGTCTCAACGGATTGCTCTATGGATTAAAGCAACAGCAG 540
Db 172 ATAATGCTCTACGCTTGGTTAGGCTTACATGAAAAAGTTGCAAAAAAACACACATGCT 231
QY 541 ATGAAGGCGCTGAATAAGATTTC 564
Db 232 AAAAAAGTTCAATCGTATGTGC 255

RESULT 11
US-09-655-908-17
; Sequence 17, Application US/09655908
; Patent No. 6645747
; GENERAL INFORMATION:
; APPLICANT: Hallahan, David L.
; TITLE OF INVENTION: cis-Prenyltransferases from Plants
; FILE REFERENCE: BC1019 US NA
; CURRENT APPLICATION NUMBER: US/09/655,908
; CURRENT FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/155,046
; PRIOR FILING DATE: 1999-09-21
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 17
; LENGTH: 1028
; TYPE: DNA
; ORGANISM: Glycine max
US-09-655-908-17

Query Match 5.8%; Score 35.4; DB 4; Length 1028;
Best Local Similarity 54.1%; Pred. No. 0.072;
Matches 72; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 398 CGCAATTCATCGCCGCAACAGCCGCACTGATGATGATATATCGTGGCGGTCAACA 457
Db 77 CCTATTCTCTCTGTTAAACACACCTTCTCCCTCTTGTATTATCTCACTATTATCA 136
QY 458 CTATTGTGGTGAATTTATTTGTGATGATCGGTTACGCCACCTTGTCTCAACGGATTGCTC 517
Db 137 CTATCGTTATCGTTATCGTTGTTATCATCTCTTCCATCACCCTTCCCAACACAGAGTCT 196
QY 518 TATGATTAAGG 530
Db 197 TATGCTCTCGAAG 209

RESULT 12
US-09-634-238-82
; Sequence 82, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J.
```

APPLICANT: Blokeberg, Leonard, N.
APPLICANT: Lubbers, Mark W.
APPLICANT: Dekker, James
APPLICANT: Christenson, Anna C.
APPLICANT: Holland, Ross
APPLICANT: O'Toole, Paul W.
APPLICANT: Reid, Julian R.
APPLICANT: Coolbear, Timothy
TITLE OF INVENTION: Polynucleotides, materials incorporating them and methods for using them.
FILE REFERENCE: 11000.1043U1
CURRENT APPLICATION NUMBER: US/09/634,238
NUMBER OF SEQ ID NOS: 422
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 82
LENGTH: 1442
TYPE: DNA
ORGANISM: Lactobacillus rhamnosus
US-09-634-238-82

Query Match
Best Local Similarity 5.3%; Score 32.8; DB 4; Length 1442;
Matches 118; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 309 TACTCAATCGGTCGACATTTTCAGGGCCAGTCTTTGTGTAATCTCACCATCCAA 368
DB 185 TAGACAGACTGCTGCATACCTTTATAGGAGAGTTAAACATCGCGTAACAAATACCC 244
QY 369 AAGTATTGTTTCTCGGCGGCTATTTCGGCAATTCATCGCGCAACAGCGCAACT 428
DB 245 AAGGACTTGTAACTCGTGCGGCACTGCTGTTACCAAGTTGAAGGGCAACCAAGAA 304
QY 429 GATGAGTATATCGTCTCGGCGTCACCACTATTGTGTCATATTATTTGATGATCGG 488
DB 305 GACGGAAGAGTTCGATGTTCTTGGGATGTTTCTGGAAGCAAGCGCGTTTAGTCCT 364
QY 489 TTACGCCACCTTGCTCAACGAGTTCCTATGATGATTAAAGCAACAGCAGATGAAGGC 548
DB 365 GACCCGCGCTGATTTTATCATGCTATGATGAGATTGCGGTAGCAGAGCATAT 424
QY 549 GCTGAATAAGATTTTCGGGT 568
DB 425 GGTCAATCAAGTAACGGCT 444

RESULT 13
US-09-252-991A-15039
Sequence 15039, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.1136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: 1999-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15039
LENGTH: 681
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15039

Query Match
Best Local Similarity 5.2%; Score 32.2; DB 4; Length 681;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 175 GGGTTGGGACGCTATTTCCGCTCAGTGATTGGTTGAAGTTGATGAGTTGGAAGTGGGACGC 234

APPLICANT: Blokeberg, Leonard, N.
APPLICANT: Lubbers, Mark W.
APPLICANT: Dekker, James
APPLICANT: Christenson, Anna C.
APPLICANT: Holland, Ross
APPLICANT: O'Toole, Paul W.
APPLICANT: Reid, Julian R.
APPLICANT: Coolbear, Timothy
TITLE OF INVENTION: Polynucleotides, materials incorporating them and methods for using them.
FILE REFERENCE: 11000.1043U1
CURRENT APPLICATION NUMBER: US/09/634,238
NUMBER OF SEQ ID NOS: 422
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 82
LENGTH: 1442
TYPE: DNA
ORGANISM: Lactobacillus rhamnosus
US-09-634-238-82

Query Match
Best Local Similarity 5.3%; Score 32.8; DB 4; Length 1442;
Matches 118; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 309 TACTCAATCGGTCGACATTTTCAGGGCCAGTCTTTGTGTAATCTCACCATCCAA 368
DB 185 TAGACAGACTGCTGCATACCTTTATAGGAGAGTTAAACATCGCGTAACAAATACCC 244
QY 369 AAGTATTGTTTCTCGGCGGCTATTTCGGCAATTCATCGCGCAACAGCGCAACT 428
DB 245 AAGGACTTGTAACTCGTGCGGCACTGCTGTTACCAAGTTGAAGGGCAACCAAGAA 304
QY 429 GATGAGTATATCGTCTCGGCGTCACCACTATTGTGTCATATTATTTGATGATCGG 488
DB 305 GACGGAAGAGTTCGATGTTCTTGGGATGTTTCTGGAAGCAAGCGCGTTTAGTCCT 364
QY 489 TTACGCCACCTTGCTCAACGAGTTCCTATGATGATTAAAGCAACAGCAGATGAAGGC 548
DB 365 GACCCGCGCTGATTTTATCATGCTATGATGAGATTGCGGTAGCAGAGCATAT 424
QY 549 GCTGAATAAGATTTTCGGGT 568
DB 425 GGTCAATCAAGTAACGGCT 444

RESULT 13
US-09-252-991A-15039
Sequence 15039, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.1136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: 1999-02-18
PRIOR FILING DATE: 1998-02-18
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15039
LENGTH: 681
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15039

Query Match
Best Local Similarity 5.2%; Score 32.2; DB 4; Length 681;
Matches 58; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 175 GGGTTGGGACGCTATTTCCGCTCAGTGATTGGTTGAAGTTGATGAGTTGGAAGTGGGACGC 234

DB 241 GGCCTGGCGGTGCTGATCACCACCACCGCTGGCTGTTCAACGGCTGAATACACCGC 300
QY 235 GCGCTTACTTGATTTGGCTGGGAATCCAGAGTGGCGGC 275
DB 301 GCGGTCTACCTGATCTGGATCGGCATCCAGGCCCTGGCTC 341

RESULT 14
US-09-641-638-278/c
Sequence 278, Application US/09641638
Patent No. 6432848
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Il'ya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET.051CF1
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 278
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 501
OTHER INFORMATION: 12-835-54 : polymorphic base A or G
NAME/KEY: misc_binding
LOCATION: 481..500
OTHER INFORMATION: 12-835-54.misl, potential
NAME/KEY: misc_binding
LOCATION: 502..521
OTHER INFORMATION: 12-835-54.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 449..468
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 879..898
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 489..513
OTHER INFORMATION: 12-835-54 potential probe
NAME/KEY: misc_feature
LOCATION: 785
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-278

Query Match
Best Local Similarity 5.2%; Score 32; DB 4; Length 1001;
Matches 74; Conservative 1; Mismatches 71; Indels 0; Gaps 0;

QY 169 TGGCTGGGTGGGAGCGCTATTTCCGCTCAGTGATTCGCTTTGAAGTTGGAAGTGG 228
DB 535 TGGGAGGGGGGGATCTACTTTTACAACTCAGCATGAGAGATGAGTGAATTT 476
QY 229 GCAGCGCGGCTTACTTGATTTGGCTGGGAATCCAGAGTGGCGCGCGCTGTGCAATT 288
DB 475 CCAGATGGTTCTCTCTGATGACTGTGGCATAATGGCAGTTTGGCCCTTGGGAGCAAT 416
QY 289 GACCTTAATCGCTGGCTCTACTCA 314

Db 415 GGGCTGCAACCCCTGAAGCCTAGACA 390

RESULT 15

US-09-634-238-195/c

; Sequence 195, Application US/09634238

; Patent No. 6544772

; GENERAL INFORMATION:

; APPLICANT: Glenn, Matthew

; APPLICANT: Havukkala, Ilkka J.

; APPLICANT: Bloksberg, Leonard, N.

; APPLICANT: Lubbers, Mark W.

; APPLICANT: Dekker, James

; APPLICANT: Christensen, Anna C.

; APPLICANT: Holland, Ross

; APPLICANT: O'Toole, Paul W.

; APPLICANT: Reid, Julian R.

; APPLICANT: Coolbear, Timothy

; TITLE OF INVENTION: Polynucleotides, materials incorporating

; FILE OF INVENTION: them and methods for using them.

; FILE REFERENCE: 11000.1043U1

; CURRENT APPLICATION NUMBER: US/09/634,238.

; CURRENT FILING DATE: 2000-08-08

; NUMBER OF SEQ ID NOS: 422

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 195

; LENGTH: 1597

; TYPE: DNA

; ORGANISM: Lactobacillus rhamnosus

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(1597)

; OTHER INFORMATION: n = A,T,C or G

US-09-634-238-195

Query Match 5.2%; Score 32; DB 4; Length 1597;
 Best Local Similarity 52.2%; Pred No. 1.4;
 Matches 71; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
 QY 361 AATCCAAAGTATTGTGTTCTGGCGCGCTATTTCGGCAATTCATCATGCCGCAACAG 420
 Db 440 AAACCGCAACGTTTACGATCGCGCTGCTTTTCGGCAGCGCGCATGGCGTTGGG 381
 QY 421 CGGCACTGATCGATATCGTCTCGCGCTCACCACCTATTGTGTCGATATTATTG 480
 Db 380 CAGCAGGGAACCGTATTTTTGTGCGATCGATCAACGGTGGGCGCTAATCGCG 321
 QY 481 ATGATCGTTACGCCA 496
 Db 320 TTGATGATGAAGCCA 305

Search completed: March 28, 2004, 18:24:34
 Job time : 68.6694 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 16:45:05 ; Search time 261.248 Seconds
(without alignments)
8764.284 Million cell updates/sec

Title: US-09-847-392-1_COPY_557_1171

Perfect score: 615

Sequence: 1 atgaccttagaatgttggtt.....tagcatcgagcgatgcg 615

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	615	100.0	1200	9	US-09-927-395-1
2	615	100.0	1200	9	US-09-847-392-1
3	414	67.3	417	12	US-10-282-122A-6496
4	414	67.3	417	14	US-10-287-374-274
5	379.2	61.7	621	12	US-10-282-122A-39598
6	375	61.0	618	12	US-10-282-122A-23973
7	357.4	58.1	618	12	US-10-282-122A-20027
8	344.8	56.1	623	12	US-10-282-122A-36679
9	274.8	44.7	621	12	US-10-282-122A-41840
10	259	42.1	259	12	US-10-282-122A-1269
11	207.4	33.7	263	14	US-10-287-374-107
12	149.8	24.4	618	12	US-10-282-122A-40780
13	112.4	18.3	630	12	US-10-282-122A-30630
14	104	16.9	630	12	US-10-282-122A-33154
15	86.4	14.0	633	12	US-10-282-122A-31699

15	80	13.0	627	12	US-10-282-122A-14918	Sequence 14918, A
17	64.6	10.5	543	12	US-10-282-122A-8689	Sequence 8689, Ap
18	58.8	9.6	633	12	US-10-282-122A-11642	Sequence 11642, A
19	42.4	6.9	536	16	US-10-338-110-119	Sequence 119, App
20	39.8	6.5	987	12	US-10-433-561-137	Sequence 137, App
21	39.8	6.5	1083	12	US-10-433-561-137	Sequence 137, App
22	39.4	6.4	1049	12	US-10-142-426-358	Sequence 358, App
23	39.4	6.4	1049	14	US-10-123-155-358	Sequence 358, App
24	39.4	6.4	1049	14	US-10-146-731-358	Sequence 358, App
25	39.4	6.4	1049	14	US-10-140-472-358	Sequence 358, App
26	39.4	6.4	1049	14	US-10-141-761-358	Sequence 358, App
27	39.4	6.4	1049	14	US-10-142-885-358	Sequence 358, App
28	39.4	6.4	1049	14	US-10-158-790-358	Sequence 358, App
29	39.4	6.4	1049	15	US-10-137-871-358	Sequence 358, App
30	39.4	6.4	1049	15	US-10-140-923-358	Sequence 358, App
31	38.4	6.4	1049	15	US-10-141-756-358	Sequence 358, App
32	39.4	6.4	1049	15	US-10-141-753-358	Sequence 358, App
33	39.4	6.4	1049	15	US-10-140-805-358	Sequence 358, App
34	39.4	6.4	1049	15	US-10-140-864-358	Sequence 358, App
35	38.4	6.2	574	14	US-10-184-644-234	Sequence 234, App
36	38.4	6.2	574	14	US-10-184-644-234	Sequence 234, App
37	36.2	5.9	686	14	US-10-184-644-524	Sequence 524, App
38	36.2	5.9	686	14	US-10-184-634-524	Sequence 524, App
39	35.6	5.8	1049	12	US-10-142-426-358	Sequence 358, App
40	35.6	5.8	1049	14	US-10-123-155-358	Sequence 358, App
41	35.6	5.8	1049	14	US-10-146-731-358	Sequence 358, App
42	35.6	5.8	1049	14	US-10-140-472-358	Sequence 358, App
43	35.6	5.8	1049	14	US-10-141-761-358	Sequence 358, App
44	35.6	5.8	1049	14	US-10-142-885-358	Sequence 358, App
45	35.6	5.8	1049	14	US-10-158-790-358	Sequence 358, App

ALIGNMENTS

RESULT 1

US-09-927-395-1
; Sequence 1, Application US/09927395
; Patent No. US20020058314A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATIEVA, NATALYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMIOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TORHAKOVA, IRINA LVONNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO LO-HOMOSERINE AND METHOD
; FOR PRODUCING L-AMINO ACIDS
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/927,395
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (557)..(1171)
US-09-927-395-1

Query Match 100.0%; Score 615; DB 9; Length 1200;
Best Local Similarity 100.0%; Pred. No. 4.3e-206;
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACCTTAGAATGGTGGTTGGCTACCTGATCGATCATTTTAAAGCTGTCGCCA 60
DB 557 ATGACCTTAGAATGGTGGTTGGCTACCTGATCGATCATTTTAAAGCTGTCGCCA 616

QY 61 GGCTCTGTCGAATCAACACTATGACCACTCGCTCGCTCAACACCGGTTATCCGGCGGTCGC 120
Db 617 GGCTCTGTCGAATCAACACTATGACCACTCGCTCGCTCAACACCGGTTATCCGGCGGTCGC 676
QY 121 GTCTATTGCTGGGCTTCAGACCGGACTCGCGATTTCATATTGCTGCTGTTGGCGTGGGGTT 180
Db 677 GTCTATTGCTGGGCTTCAGACCGGACTCGCGATTTCATATTGCTGCTGTTGGCGTGGGGTT 736
QY 181 GGGACGCTATTTCCGCTCAGTGATTCGTTGAGTGTGAAAGTGGGCGGCGGCT 240
Db 737 GGGACGCTATTTCCGCTCAGTGATTCGTTGAGTGTGAAAGTGGGCGGCGGCT 796
QY 241 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCATTCATCGCCGCAACAG 300
Db 797 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCATTCATCGCCGCAACAG 856
QY 301 CTGGCCCTTACTCAATCGGCTCGCATTTGTTTCAGCGCGCAGTTTGTGAATCTCACC 360
Db 857 CTGGCCCTTACTCAATCGGCTCGCATTTGTTTCAGCGCGCAGTTTGTGAATCTCACC 916
QY 361 AATCCCAAAAGTATTGTTCTCGCGCGCTATTTCGGCAATTCATCGCCGCAACAG 420
Db 917 AATCCCAAAAGTATTGTTCTCGCGCGCTATTTCGGCAATTCATCGCCGCAACAG 976
QY 421 CGCAACTGATGAGTATATGCTGCTCGCGCTCACCATTGTTGCTGCTGATATTATGTG 480
Db 977 CGCAACTGATGAGTATATGCTGCTCGCGCTCACCATTGTTGCTGCTGATATTATGTG 1036
QY 481 ATGATCGGTTAGCCACCCCTTCTCAACGGATTGCTCATGATTAAGGACCAACAG 540
Db 1037 ATGATCGGTTAGCCACCCCTTCTCAACGGATTGCTCATGATTAAGGACCAACAG 1096
QY 541 ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGGAGCGCTGTAGCA 600
Db 1097 ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGGAGCGCTGTAGCA 1156
QY 601 TCGGCGAGGCAATGCG 615
Db 1157 TCGGCGAGGCAATGCG 1171

RESULT 2

US-09-847-392-1
; Sequence 1, Application US/09847392
; Patent No. US20020102670A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VERNAMIOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHMAKOVA, IRINA LVONNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US/09/847,392
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (557)..(1171)
US-09-847-392-1

Query Match 100.0%; Score 615; DB 9; Length 1200;
Best Local Similarity 100.0%; Pred. No. 4.3e-206; Indels 0; Gaps 0;
Matches 615; Conservative 0; Mismatches 0;
QY 1 ATGACCTTAGAATCGTGGTTTGGCTTACCTGACATCGATCATCTTTAAAGCTGTGCGCA 60
Db 557 ATGACCTTAGAATCGTGGTTTGGCTTACCTGACATCGATCATCTTTAAAGCTGTGCGCA 616
QY 61 GGCTCTGTCGAATCAACACTATGACCACTCGCTCAACACCGGTTATCCGGCGGTCGC 120
Db 617 GGCTCTGTCGAATCAACACTATGACCACTCGCTCAACACCGGTTATCCGGCGGTCGC 676
QY 121 GTCTATTGCTGGGCTTCAGACCGGACTCGCGATTTCATATTGCTGCTGTTGGCGTGGGGTT 180
Db 677 GTCTATTGCTGGGCTTCAGACCGGACTCGCGATTTCATATTGCTGCTGTTGGCGTGGGGTT 736
QY 181 GGGACGCTATTTCCGCTCAGTGATTCGTTGAGTGTGAAAGTGGGCGGCGGCT 240
Db 737 GGGACGCTATTTCCGCTCAGTGATTCGTTGAGTGTGAAAGTGGGCGGCGGCT 796
QY 241 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCATTCATCGCCGCAACAG 300
Db 797 TACTTGATTGGCTGGGAATCCAGCAGTGGCGCGCTGGTGCATTCATCGCCGCAACAG 856
QY 301 CTGGCCCTTACTCAATCGGCTCGCATTTGTTTCAGCGCGCAGTTTGTGAATCTCACC 360
Db 857 CTGGCCCTTACTCAATCGGCTCGCATTTGTTTCAGCGCGCAGTTTGTGAATCTCACC 916
QY 361 AATCCCAAAAGTATTGTTCTCGCGCGCTATTTCGGCAATTCATCGCCGCAACAG 420
Db 917 AATCCCAAAAGTATTGTTCTCGCGCGCTATTTCGGCAATTCATCGCCGCAACAG 976
QY 421 CGCAACTGATGAGTATATGCTGCTCGCGCTCACCATTGTTGCTGCTGATATTATGTG 480
Db 977 CGCAACTGATGAGTATATGCTGCTCGCGCTCACCATTGTTGCTGCTGATATTATGTG 1036
QY 481 ATGATCGGTTAGCCACCCCTTCTCAACGGATTGCTCATGATTAAGGACCAACAG 540
Db 1037 ATGATCGGTTAGCCACCCCTTCTCAACGGATTGCTCATGATTAAGGACCAACAG 1096
QY 541 ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGGAGCGCTGTAGCA 600
Db 1097 ATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGGGAGCGCTGTAGCA 1156
QY 601 TCGGCGAGGCAATGCG 615
Db 1157 TCGGCGAGGCAATGCG 1171

RESULT 3

US-10-282-122A-6496
; Sequence 6496, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Chlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining prior application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: Patent in version 3.1
;; SEQ ID NO 6496
;; LENGTH: 417
;; TYPE: DNA
;; ORGANISM: Escherichia coli
US-10-282-122A-6496

Query Match 67.3%; Score 414; DB 12; Length 417;
Best Local Similarity 100.0%; Pred. No. 2.5e-135;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 202 GTGATGGCTTTGAAGTCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 261
DB 1 GTGATGGCTTTGAAGTCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 60
QY 262 CAGCAGTGGCGGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 321
DB 51 CAGCAGTGGCGGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 120
QY 322 CGACATTTGTTCCAGCGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 381
DB 121 CGACATTTGTTCCAGCGGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 180
QY 382 CTGGCGGCGGCTTATTCGCAATTCATCATCGGCAACAGCGGCAACTGATGCGATATATC 441
DB 181 CTGGCGGCGGCTTATTCGCAATTCATCATCGGCAACAGCGGCAACTGATGCGATATATC 240
QY 442 GTGCTCGGCGTCACCACTATTGTGTCGATATTTATGTCGATCGGTTACGCCACCCCTT 501
DB 241 GTGCTCGGCGTCACCACTATTGTGTCGATATTTATGTCGATCGGTTACGCCACCCCTT 300
QY 502 GCTCAACGGATTGCTCTATGATTAAGGACCAAGCAGATGAAGCGCTGAATAGATT 561
DB 301 GCTCAACGGATTGCTCTATGATTAAGGACCAAGCAGATGAAGCGCTGAATAGATT 360
QY 562 TTCGGCTCGTGTGTTATCTGCTGGGAGCGCTGTAGCATCGGCGAGGCGATGCG 615
DB 361 TTCGGCTCGTGTGTTATCTGCTGGGAGCGCTGTAGCATCGGCGAGGCGATGCG 414

RESULT 4
US-10-287-274-274
; Sequence 274, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THEREOF
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09

;; PRIOR APPLICATION NUMBER: US 09/711164
;; PRIOR FILING DATE: 2000-11-09
;; NUMBER OF SEQ ID NOS: 469
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 274
;; LENGTH: 417
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(417)
US-10-287-274-274

Query Match 67.3%; Score 414; DB 14; Length 417;
Best Local Similarity 100.0%; Pred. No. 2.5e-135;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 202 GTGATTCGGTTGAAGTCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 261
DB 1 GTGATTCGGTTGAAGTCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 60
QY 262 CAGCAGTGGCGGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 321
DB 61 CAGCAGTGGCGGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 120
QY 322 CGACATTTGTTCCAGCGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 381
DB 121 CGACATTTGTTCCAGCGCGGCTTGAAGTGGCGGCGGCTTACTTGTGTTGGCTGGGAATC 180
QY 382 CTGGCGGCGGCTTATTCGCAATTCATCATCGGCAACAGCGGCAACTGATGCGATATATC 441
DB 181 CTGGCGGCGGCTTATTCGCAATTCATCATCGGCAACAGCGGCAACTGATGCGATATATC 240
QY 442 GTGCTCGGCGTCACCACTATTGTGTCGATATTTATGTCGATCGGTTACGCCACCCCTT 501
DB 241 GTGCTCGGCGTCACCACTATTGTGTCGATATTTATGTCGATCGGTTACGCCACCCCTT 300
QY 502 GCTCAACGGATTGCTCTATGATTAAGGACCAAGCAGATGAAGCGCTGAATAGATT 561
DB 301 GCTCAACGGATTGCTCTATGATTAAGGACCAAGCAGATGAAGCGCTGAATAGATT 360
QY 562 TTCGGCTCGTGTGTTATCTGCTGGGAGCGCTGTAGCATCGGCGAGGCGATGCG 615
DB 361 TTCGGCTCGTGTGTTATCTGCTGGGAGCGCTGTAGCATCGGCGAGGCGATGCG 414

RESULT 5
US-10-282-122A-39598
; Sequence 39598, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26

```

; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 39598
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Salmonella typhi
US-10-282-122A-39598

Query Match          61.7%; Score 379.2; DB 12; Length 621;
Best Local Similarity 78.0%; Pred. No. 6.2e-123;
Matches 482; Conservative 0; Mismatches 133; Indels 3; Gaps 2;

QY 1 ATGACCTTAGAATGGTGGTTTCCTACCTGCTGATCGATCATTTTAACGCTTCGCCA 60
DB 1 ATGACCTTTGAATGGTGGTTTCCTATCTACTGACCTCAACTCTGTGTGAGTCTTTCGCG 60
QY 61 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACCAAGGTTATCCGCGCG--GTG 118
DB 61 GGTTCAGGCGCCATCAATACCAGCGCTCTATCAACCATGATATCGTGGCGCAGCG 120
QY 119 GCCTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATGCTGTGGCTGGGG 178
DB 121 GCTTCATCGCGGGCTCCAGACCGGGCTGGGGATACATATCGTACTGGTGGCGCTCGGA 180
QY 179 TTGGG-ACGCTATTTTCCGCGCTCAGTGATGCTGTTGAAGTGTGAAGTGGCGGCGCG 237
DB 181 CTGGGTACGCTCTTTTCGCGCTCGCTCATCGCTTTTGAATTCGAAATGCGCTGGCGCG 240
QY 238 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGTGTGCAATTGACCTTAAA 297
DB 241 GCTTATCTTATCTGGCTGGGTATCCAGCAATGGCGCGCGCGCTATCGATCTGCAT 300
QY 298 TCCTCGGCTCTACTCAATCGCGTTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTC 357
DB 301 ACTCTCGCCAGAGCAATTCGCGGGTGGCTGTTCAACGGGCGATATTTGTCAATCTA 360
QY 358 ACCAATCCCAAAAGATTTGTGTTCTGGCGGCGCTATTTCCGCAATTCATATCGCGCAA 417
DB 361 ACCAATCCCAAAAGATTTGTGTTCTTTCGCGCGCTGTTTCGCGCAATTCATATCGCGCAG 420
QY 418 CAGCCGCAACTGATGACGATATATGCTGCTCGCGCTCACCATATTTGCTGATATTTAT 477
DB 421 CAACCGCACTGGCGAGTACTCTATTTCTGGGCTCACACGATTTGGTGGATATGAT 480
QY 478 GTGATGATCGGTATCGCCACCTTCTCTCAACGATTTGCTCTATGATTAAGAGCAACAAAG 537
DB 481 GTGATGACCGGTTATGCCACGCTGGCGAGCGATTCGCGCTGGATTAAGAGCAACAAAG 540
QY 538 CAGATGAAGCGCTGAATAAGATTTTCGGCTCGTTGTTTATGCTGTGGAGCGCTGTTA 597
DB 541 CAGATGAAGCGCTGATTAAGCGTTTGTGTTGTTTATGCTGTGGAGCGCTCCTG 600
QY 598 GCATCGCGGAGCATGCG 615
DB 601 GCGTCGCAAGCACGCG 618
```

RESULT 6

```

US-10-282-122A-23973
; Sequence 23973, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 23973
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23973
```

```

Query Match          61.0%; Score 375; DB 12; Length 618;
Best Local Similarity 77.6%; Pred. No. 1.9e-121;
Matches 479; Conservative 0; Mismatches 135; Indels 3; Gaps 2;

QY 1 ATGACCTTAGAATGGTGGTTTGCCTACCTGCTGACATCGATCATTTTAACGCTTCGCCA 60
DB 1 ATGACCATTTGATGGTGGTTCGCTACCTGCTGACATCAATTTCTCAGCCTGTGCGCG 60
QY 61 GGCTCTGTGCAATCAACACTATGACCACTCGCTCAACACCGTTA--TCCGGCCGGTG 118
DB 61 GGTTCGAGCGGATTAATACCATGACCACCCTCCATTAAACACGATACCGCGCGCGCG 120
QY 119 GCGTCTATTGCTGGGCTTCAGACCGGACTGCGGATTCATATTGCTGTGGCG-TGGG 177
DB 121 GCGTCTGATTTCCGCGCTTCAGACCGGCTGGTTATTCATATCGTGTGCTGTGGCTTGGT 180
QY 178 GTTGGAGCGCTATTTTCCGCTCAGTGATTCGCTTTGAAGTGTGAAGTGGGCGCGCG 237
DB 181 CTCGCACTCTCTTCTCCCGCTCGGTGCTGGCTTTGAGGTGCTGAAATGGGCGCGCGC 240
QY 238 GCTTACTTGATTTGGCTGGGAATCCAGCAGTGGCGCGCGCTGCTCAATTGACCTTAA 297
DB 241 GCCTACCTGATCTGGCTGGGTATCCAGCAATGGCGGCGCGCGGATTCGACCTGAAT 300
```

QY 298 TCCTGGCTCTACTCAATCGGTCGACATTTGTTCCAGCGCGCAGTTTGTGTAATCTC 357
Db |||||
QY 301 ACCTGGCGAAGCGCAGACGGCGGCAAACTGTTCCAGCGCGGCTGTGTTAAATCTC 360
Db |||||
QY 358 ACCAATCCCAAAAGATTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGCA 417
Db |||||
QY 361 ACCAATCCCAAAAGATTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGCA 420
Db |||||
QY 418 CAGCCCAACTGATGACGATATCGTGTCTGGCGTCACCACTATTGTGGTCGATATTAT 477
Db |||||
QY 421 CAGCGCAGGTGATGCAATATCTCGTTCTGGCGTCACCACTCGTCGATCAT 480
Db |||||
QY 478 GTGATGATCGGTTACCGCACCTTCTGTCACGATGCTCTATGATTAAGGACCAAG 537
Db |||||
QY 481 GTGATGATCGGTTACCGCACCTTCTGTCACGATGCTCTATGATTAAGGACCAAG 540
Db |||||
QY 538 CAGATGAAGCGCTGTAATGAATTTTCGGCTCGTTGTTTATGCTGTTGATGCTGCTC 597
Db |||||
QY 541 CAGATGAAGCGCTGTAATGAATTTTCGGCTCGTTGTTTATGCTGTTGATGCTGCTC 600
Db |||||
QY 598 GCATCGCGGCGCATGC 614
Db |||||
QY 601 GCCTCCGCGCGCATGC 617
Db |||||

RESULT 7

US-10-282-122A-20027
; Sequence 20027, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20027
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Enterobacter cloacae

US-10-282-122A-20027

Query Match 58.1%; Score 357.4; DB 12; Length 618;
Best Local Similarity 75.9%; Pred. No. 3.2e-115;
Matches 468; Conservative 0; Mismatches 146; Indels 3; Gaps 2;
QY 1 ATGACCTTAGAATGGTGGTTTGGCTTACCTGCTGACATCGATCATTTTAAAGCTGTGCCA 60
Db 1 ATGACCTTCGAATGGTGGTTTGGCTTACCTGCTGACATCGATCATTTTAAAGCTGTGCCG 60
QY 61 GCCTCTGGTGCATCAACACTATGACGACCTCGCTCGCTCAACACGGTTA--TCCGCCCGGTG 118
Db 61 GCCTCGGGGGCTATTAAACCATGACCACTCCATCAATACGGCTACCGCGGTGCGGCG 120
QY 119 GGGCTATTGCTGGGCTTTCAGACCGGACTGCGGATTCATATTGCTGCTGGTGGGCT--GGG 177
Db 121 GCATCGATTCGGTTTCAGACCGGCTGGGATTCATATTGCTGCTGGGATCGGT 180
QY 178 GTTGGAGCGCTATTTTCCGGCTCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 237
Db 181 CTGGGGACCCCTGTTCTCCCGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 238 GCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGCGCGCTGCTGCTGCTGCTGCTGCTGCT 297
Db 241 GCGTATCTGATTTGGCTGGCATCCAGCAGTGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 298 TCGCTGGCTCTACTCAATCGCTCGCATTTGTTTCAGCGCGCAGTTTGTGAAATCTC 357
Db 301 ACGCTGGCTCTGACGCAAAACCGCGGCTCTGTTTAAAGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 358 ACCAATCCCAAAAGATTGTTTCTGGCGGCTGTTTCCGCAATTCATCATCGCGCAA 417
Db 361 ACCAATCCCAAAAGATTGTTTCTGGCGGCTGTTTCCGCAATTCATCATCGCGCAT 420
QY 418 CAGCGCAACTGATGACGATATGCTGCTCGGCTCACCACTATTGTGGTCGATATTAT 477
Db 421 CAGCTCAGGTGATGACGATGCTGCTGCTGGCGCGCAGCTATTATTCGTCGATCATC 480
QY 478 GTGATGATCGGTTACGCCACCTTCTCAACGGATTGCTCTATGATTAAGGACCAAG 537
Db 481 GTGATGATGTTTACGCCACCTTCTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 538 CAGATGAAGCGCTGAATGAATTTTCGGCTCGTTGTTTATGCTGCTGGAGCGCTGTTA 597
Db 541 CAGATGAAGCGCTGAATGAATTTTCGGCTCGTTGTTTATGCTGCTGGAGCGCTGTTA 600
QY 598 GCATCGCGGCGCATGC 614
Db 601 GCGTCAGCGCGTCACGC 617
Db |||||

RESULT 8

US-10-282-122A-36679
; Sequence 36679, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 36679
;; LENGTH: 623
;; TYPE: DNA
;; ORGANISM: Salmonella paratyphi A
US-10-282-122A-36679

Query Match 56.1%; Score 344.8; DB 12; Length 623;
Best Local Similarity 76.8%; Pred. No. 98-111;
Matches 475; Conservative 0; Mismatches 137; Indels 8; Gaps 4;
Qy 1 ATGACCTTAGAATGGTGGTTGGCTACCTGCTGACATCGATCATTTAAACGCTGCGCA 60
Db 1 ATGACCTTAGAATGGTGGTTGGCTATCTACTGACCTCACTCTGCTGAGTCTTCTCCG 60
Qy 61 GGCTCTGGTCAATCAACATATGACCACTCGCTCAACACCGGTTATCCGGCG -GTG 118
Db 61 GGTTAGGGGCCCAATCAATACGACGCTATCAACCATGATATCGTGGCGGCG 120
Qy 119 GCCTCTATTGCTGGCTTCAGACCGGACTGGGATTCATATTTGCTGTTGGGTGGG 178
Db 121 GCTTCTATCGCGGACTCAGACCGGACTGGGATACATATCTGCTGCTGGTGG 180
Qy 179 TTGGG-AGCTATTTCCGCTCAGTATGCTGTTGAAGTGTGAAGTGGCGAGCGCG 237
Db 181 CTGGGTACGCTCTTTTCGGGCTCGCTCTCGCTTTTGAATTCGAAATGGCTGGCG 240
Qy 238 GCTTACTTGTGTTGGTGGGATCCAGACGCTGGCGCGCTGGTCAATGACTTAA 297
Db 241 GCTTATCTTCTGCTGGGTATCCAGCAATGGCG---CGAGGGCTATCGATCTGCAT 297
Qy 298 TCCTGGGCTCTACTCAATCGGCTGACATTTGTTCCAGCGCGAGTGTGTTGTAATCTC 357
Db 298 ACTCTGCGCCAGACGCAATCGCGGGTGGCTGTTTCAACGGGCGATATTTGTAATCTA 357
Qy 358 ACCAATCCCAAGATGTTGTTCTGGCGGGCTATTTCCGCAATTCATCATCCGCA 417
Db 358 ACCAATCCCAAGATGTTGTTCTTTCGCGGCTGTTTCGCGCAATTCATCATCCGCG 417
Qy 418 CAGCCGCAACTGATGAGTATATCGTGTGCGGCTCAGCACTATTTGTTGGTGCATATTAT 477
Db 418 CAACCGCAACTGCGCGAGTACCTCATTTCTCGGCTCAGCACTGTTGGTGGATGAT 477
Qy 478 GTGATATGTTTACCCACCTTGTCTCAACGG--ATTGCTCTATGATTAAGACCAA 535
Db 478 GTGATACCGGTTTACCCACACTGGCGCTAGCGCACTTCCGCGCTGGATTAAGACCAA 537
Qy 536 AGCAGATGAAGCGCTGAATAAGATTTTCGGCTCGTGTGTTATGCTGTGGGAGCGCTGT 595
Db 538 AGCAGATGAAGCGCTGAATAAGCGTTTGGTTCGTTTATCTGTTAGCGCGCTCC 597
Qy 596 TACATCGGGAGGCAATGCG 615

Db 598 TGGCGTCGCGACAGACGCG 617
RESULT 9
US-10-282-122A-41840
;; Sequence 41840, Application US/10282122A
;; Publication No. US20040029129A1
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Liangsu
;; APPLICANT: Zamudio, Carlos
;; APPLICANT: Malone, Cheryl
;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Kari
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John
;; APPLICANT: Carr, Grant
;; APPLICANT: Yamamoto, Robert
;; APPLICANT: Forsyth, R.
;; APPLICANT: Xu, H.
;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 41840
;; LENGTH: 621
;; TYPE: DNA
;; ORGANISM: Yersinia pestis
US-10-282-122A-41840

Query Match 44.7%; Score 274.8; DB 12; Length 621;
Best Local Similarity 67.5%; Pred. No. 4.9e-86;
Matches 416; Conservative 0; Mismatches 197; Indels 3; Gaps 2;
Qy 1 ATGACCTTAGAATGGTGGTTGGCTACCTGCTGACATCGATCATTTTAAACGCTGCGCA 60
Db 1 ATGACCTTAGAATGGTGGTTAACTATCTGCTGACAACTGATCTCTCTCC 60
Qy 61 GGCTCTGGTCAATCAACACTATGACCACTCGCTCAACCAAGTTATC--CGCCCGGTG 118
Db 61 GGCTCGGGGGCCATTAAACAAATGAGCACTGCCATGACCCGTGGGGTGT 120
Qy 119 GGCTCTATTGCTGGCTTCAGACCGGACTGGCGATTCATATTTGCTGTGGCGT-GGG 177
Db 121 GCTTCCATTGCGGGTTACAACTGGGCTGCGGTGCATATCGTCTTGTGTTGGCTCGGT 180
Qy 178 GTTGGAGCGCTATTTTCCCGCTCAGTATTCGCTTTCAGCTGTTGAAGTGGCGAGCGCG 237
Db 181 CTGTGTCGCCCTGGTCTCTCAATCTTACTGGCATTTGAAATATTAAGTGGCTCGGTGG 240

```
QY 238 GCTTACTTGTGCTGGGAATCCAGCAGTGGCGCGCTGGTGCATTTGCACTTAAA 297
Db 241 GCTTACTTAAATTTGGCTTGTGTTTCAACAATGCGTGCAGCGCGTGCCTTGCCTGCAT 300
QY 298 TCCTGGCTCTACTCAATTCGCGTGCACATTTGTTCCAGCGCGCAGTTTGTGGAATCTC 357
Db 301 GCACTGGCAAAACAGTATGCCAGGCGTAAGCTGTTTAAACGTGGGTGTTGTTAACCTG 360
QY 358 ACCAATCCCAAAAGATTGTTTCTGGCGGCTATTTCGCGCAATTCATCATGCCGCAA 417
Db 361 ACCAATCCCAAAAGCAATTTTCTGGCGGTTATTTCACCAATTTGACTGCCACAG 420
QY 418 CAGCGCAACTGATGCAGTATATCTGCTCGCGCTCACCACATTTATGTTGGTGCATATTATT 477
Db 421 CAACCCAGGTGGCACAGTATTGATTTTGGCGACACACAGCTGATGTTGCGATATTATC 480
QY 478 GTGATGATCGTTAGCCACCCCTTCTCAACGGATTGCTTATGATTAAGACCAAAG 537
Db 481 GTGATGATTTGGTTATGCCACTTTGGCTACAGTATGCGCGCTTGGATTAATCGCACAG 540
QY 538 CAGATGAAGCGCGCTGAATAAGATTTTCGCGCTGTTGTTTATGCTGGTGGAGCGCTGTTA 597
Db 541 CAAATGAACCTTTGAACCGAATATTTCGCGCGCTATTATTGTTGATTTGATTTGCAATTATTA 600
QY 598 GCATCGCGCGAGCATG 613
Db 601 GCCACAGCGCGCAAAG 616
```

RESULT 10

```
US-10-282-122A-1269/c
; Sequence 1269, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1269
; LENGTH: 259
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-282-122A-1269
```

Query Match 42.1%; Score 259; DB 12; Length 259;

Best Local Similarity 100.0%; Pred. No. 1.1e-80;

Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 CTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCACCAATCCCA 367

Db 259 CTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCACCAATCCCA 200

QY 368 AAGATTTATGTTCTTCCGCGCGCTATTTCGCAATTCATCATGCCGCAAC 427

Db 199 AAGATTTATGTTCTTCCGCGCGCTATTTCGCAATTCATCATGCCGCAAC 140

QY 428 TGATGCAGTATATCGTCTGGCGTCAACACTATTGTTGGTGCATATTATTGTCATGATCG 487

Db 139 TGATGCAGTATATCGTCTGGCGTCAACACTATTGTTGGTGCATATTATTGTCATGATCG 80

QY 488 GTTACGCCACCCCTTCTCAACGGATTGCTCTATGGATTAAAGCAACCAAGCAGATGAAG 547

Db 79 GTTACGCCACCCCTTCTCAACGGATTGCTCTATGGATTAAAGCAACCAAGCAGATGAAG 20

QY 548 CGCTGAATAAGATTTTCGG 566

Db 19 CGCTGAATAAGATTTTCGG 1

RESULT 11

```
US-10-287-274-107/c
; Sequence 107, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107
; LENGTH: 263
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-287-274-107
```

Query Match 33.7%; Score 207.4; DB 14; Length 263;

Best Local Similarity 94.3%; Pred. No. 2e-62;

Matches 248; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 305 CCTCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCACCAATC 364

Db 263 CCCTACTCAATCGCTCGACATTTGTTCCAGCGCGCAGTTTGTGAATCTCACCACCAATC 204

QY 365 CCAAAAGTATTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCGCG 424

Db 203 CCAAAAGTATTGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCCGCAACAGCGCG 144

QY 425 AACTGATGCAGTATATCGTCTC-GCGGTCACCACTATTGTTGGTGCATATTATTGTCATG 483

Db 143 AACTGATGCAGTATATCGTCTC-GCGGTCACCACTATTGTTGGTGCATATTATTGTCATG 84

QY 484 ATC-GGTTAGCCACCCCTTGTCTCAACGGATTG--CTCTATGGATTAAAGCAACCAAGCAG 540

```
Db 83 ATCGGTTACGCCACCCCTTGCCTAACCGGATGGCTCTTAAGGATTAAGAGCCCAACAA 24
QY 541 ATGAAGCGCTGAATAAGATTTT 563
Db 23 ATGAAGCGCTGAATAAGATTTT 1
RESULT 12
US-10-282-122A-40780
; Sequence 40780, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40780
; LENGTH: 618
; TYPE: DNA
; ORGANISM: Vibrio cholerae
US-10-282-122A-40780
Query Match 24.4%; Score 149.8; DB 12; Length 618;
Best Local Similarity 55.5%; Pred. No. 7.3e-42;
Matches 330; Conservative 0; Mismatches 262; Indels 3; Gaps 2;
QY 16 TGGTTTGCTACTCGTCGATCGATCATTTTAAAGCTGTGCGCAGGCTCTGGTGAATC 75
Db 16 TGGCTTGCTATCTACTACCGCGGTGTGTAGTCTGGCGCCCGTTTACGACGGTT 75
QY 76 AACATATGACACCTCGCTCAACACCGGT-TATCCGGCGGTGGCGTCTATTGTGGG 133
Db 76 AACTCGATCAGCAATGGTTTAAAGTTATGCTACAGCCACTCATTAGGGCGGATTCGGC 135
QY 134 CTTTCAGACCGGACTGCGGATTCATATTGCTGTGCTGCG-CTGGGTTGGAGCTATT 192
Db 136 CTACAAATCGGTCTTGCTTGCTATAGTATGTTGGCATTTGAATGGCTTTGGTG 195
```

```
QY 193 TCCCGCTCAGTGATTTGGCTTTTGAAGTGTTCAGATGGGCGAGCGGCTTACTTGAATTGG 252
Db 196 GCACAATCTGGTTGGCGTTCACTCTGATCAAAATGGATTGGCGCGCTTATTTAGTATGG 255
QY 253 CTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTCACCTTAAATCGCTGGCTCTACT 312
Db 256 CTGGGTATTTCAGAAATGGCGAGATAGAGCCCCCTAACGGCTACCAACAACCTTCTCATGAA 315
QY 313 CAATCGCGTCGACATTTGTTTCCAGCGCGCAGTTTGTGAATCTCCACCAATCCCAAAAGT 372
Db 316 CTCTCACAAGCGCCCTGCTGCGTAAAGCAGTAGTACTGATTAACTGACCAATCCCAATCA 375
QY 373 ATTGTGTTTCTGCGCGCGCTATTTCGCCAATTCATATGCGGCAACAGCGCGAAGTATG 432
Db 376 ATCGTTTCTTTAGTGGCATTTATTTCCGCACTTTATTGATCCCACTAGAGATCATTTGCCG 435
QY 433 CAGTATATCGTGTGCGCGCTCACCACTATTGTGTGATATATTGTGATGATCGGTATC 492
Db 436 CAGTTCITGGTTCCTGGGTATCACACCGTTACTATTGATCGGATTGTCTATTCGGATAC 495
QY 493 GCCACCTTGCTCAACGGATTGCTCTATGGATTAAAGGACCAAGCAGATGAAGGGCGCTG 552
Db 496 ACGGCACTGGCTGCGCAGCTAGTGCCTATATTTCGTTCACTAACATAATGACCAAGATG 555
QY 553 AATAAGATTTTCGGCTCGTTGTTTATGCTGTGGAGCGCTGTAGCATCGGCGA 607
Db 556 AATAAATCTGTTGGTTCGATGTTTATGCGCTGCGGATGCTGTGCTACCGCCA 610
```

```
RESULT 13
US-10-282-122A-30630
; Sequence 30630, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30630
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30630

Query Match 18.3%; Score 112.4; DB 12; Length 630;
Best Local Similarity 53.5%; Pred. No. 1.2e-28;
Matches 324; Conservative 0; Mismatches 276; Indels 6; Gaps 4;

QY 16 TGGTTTGCTACCTGCTGACATGATCAATTTAAAGCTGTGCGCAGGCTCTGTGTGAATC 75
DB 16 TGGTTTGGCTCTTCTCGCTGTGGCCATCAGCTGTGCGCGGCGCGGAGCATC 75
QY 76 AACATATGACCATCTGCTCAACACGGTTATCGGCCGTGTC--GTCTATTGTGGG 133
DB 76 TGGTTTGGCTCTTCTCGCTGTGGCCATCAGCTGTGCGCGGCGCGGAGCATC 75
QY 134 CTTACAGACCGGACTGGCGATTCAATTTGCTGTGGCG-TGGGGTTGGAGCGCTATTT 192
DB 136 CTGCAGATCGGCTGCGCTGCAGATCCCAATTTGCGCGCGCGCTCGTGGTGTCTG 195
QY 193 TCCGCTGATGATGCTGTTGAAGTGTGAAGTGGGAGCGCGGCTTACTTGAATTTGG 252
DB 196 GCGACTCGGCACTGGCTTTTCCAGCTGATCAAGTGGTTTGGCGGTGGCTACCTGGTATC 255
QY 253 CTGGGAATCCAGCAGTGGC--GCGCGCTGTGCAATTTGACCTTAAATCGCTGCTCTTA 310
DB 256 CTGGCGGTGGCCAGTGGCAGCGCGCGCCACAGCGCTTGACCGCATGGGAGCGCT 315
QY 311 CTCATATCG-CGTGACATTTGTTCCAGCGCGCATTTTGTGATCTCAACATCCCAA 369
DB 316 CTGGGCGGACCGTTGACCTGCTGTGCTGTTTCTGCTGATCAAGCGCAATCCCAAG 375
QY 370 AGTATTGTTTCTGGCGCGCTATTTCCGCAATTCATGCGCAACAGCGCGCAACTG 429
DB 376 GCGGTGATCTTCATGCTCGCGTGTGCGCGATTCATGACCGCGACCGCGCTGCTG 435
QY 430 ATGAGTATATGCTGCTCGCGTCAACCACTATGCTGTCGATATTAATGATGATCGGT 489
DB 436 GCGCAATACCTGATCATGGCGGCGACCATGATCTGCTGACCTGATGCTCATGGCGCG 495
QY 490 TACGCCACCTTGTCAACGATTCCTCTATGATTAAGGACCAAGCAGATGAAGCGG 549
DB 496 TACACCGGCTGCTGCGCGGTGCTAGCGTACTGCGTTGCGCGCGCGCGAGAGCTG 555
QY 550 CTGAATAAGATTTTGGCTGCTGTTGTTATGCTGTGGAGCGCTGTTAGCATCGCGAGG 609
DB 556 GTGAACCGTACCTTCGCGAGCTGTTCGTGCGTGGCGGCGCTGCTGGCGAGTACGC 615
QY 610 CATGCG 615
DB 616 CGAGCG 621

RESULT 14
US-10-282-122A-33154

; Sequence 33154, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33154
; LENGTH: 630
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-282-122A-33154

Query Match 16.9%; Score 104; DB 12; Length 630;
Best Local Similarity 54.3%; Pred. No. 1.1e-25;

Matches 278; Conservative 0; Mismatches 225; Indels 9; Gaps 3;

QY 1 ATGACCTTAGAATGGTGGTTTGGCTTACCTGCTGACATCGATCAATTTTAAAGCTGTGCGCA 60
DB 1 ATGGCACTTGAACAGTGGCTGGGCTTTTTCGGCGCTGCTGGATTATCAGCCTCTCTCT 60
QY 61 GGCTCTGTGTCATCAACACATATGACCCACTCGCTCAACACACGGTTATCGGCGCGTGGC 120
DB 61 GGCGGGGTGCCATCGCTCGATGTCCTGGCGCTGCAATACGTTTCTGCGGCTAC 120
QY 121 GTCATTGCT--GGGCTTACAGCGGACTGGCGATTCAATTTGCTGCTGCTGCGTGGG 178
DB 121 TGGAAATGCTTGGGCTGTCAGATAGCGCTGGTGGCGCAGATCGGATTGTCGCGCGCGG 180
QY 179 -TTGGAGCGTATTTCCCGCTCAGTGATGCTGTTTGAAGTGTGAAGTGGGCGCGCG 237
DB 181 CTTGGCGCGTATTTGGCAGCTTCGAGATGGCGCTTCACGCTGATCAAGTGGTTTCGCGGTG 240
QY 238 GCTTACTTGGTTGGTGGGAATCCAGCAGTGGCGG-----CCGCTGGTGAATTCAT 291
DB 241 GCGTATCTGCTATCTGCGCATCAACACAGTGGCGCGCATCGCCACCGACCTTGGCGGAC 300
QY 292 CTTAAATCGCTGCGCTTACTCAATCGCTGCGCATTTGTTCCAGCGCGCAGTTTGTG 351
DB 301 GAATCCCGCGTGGCGCGGTGGGCAAGCGGATGACGCTGTTTCGCGCGGTTTCTGGTTC 360
QY 352 AATCTCAACCAATCCAAAAGTATTGTTTCTGGCGCGCTATTTCCGCAATTCATCATG 411
DB 361 AACATCAGCAACCCCAAGCGCTGATCTTCATCTCGCGATCCTGCGCAGTTCATCGAG 420
QY 412 CCGCAACAGCGCAACTGATGTCAGTATATCGTCTCGGCGTACCACTATTGTGTCGAT 471
DB 421 CCGACTGCTCGCTGTTTCATGTCAGTACGTGATCATCGCGCGACCATGCTGCTGTCGAT 480
QY 472 ATTATTGATGATCGGTTACGCCACCGCTTGC 503
DB 481 CTGATCGTATGGCGGTTACAGCGGCTGGC 512

RESULT 15

US-10-282-122A-31699
; Sequence 31699, Application US/10282122A
; Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlson, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 31699

LENGTH: 633

TYPE: DNA

ORGANISM: Pseudomonas putida

US-10-282-122A-31699

Query Match 14.0%; Score 86.4; DB 12; Length 633;
Best Local Similarity 49.5%; Pred. No. 1.9e-19;
Matches 309; Conservative 0; Mismatches 306; Indels 9; Gaps 3;

QY	1	ATGACCTTAGAATGGTGGTTTGCTACCTGCTGATCGATCGATCATTTTAAACGCTGTCGCCA	60
DB	4	ATGTCGATGGAATATGGTTAGGCTTTTTCGCCGCTGCTGGGTGATCAGCCTTTACCC	63
QY	61	GGCTCTGGTGAATCAACTATGACACTACCTGCTCAACACCGGTTATCCGGCCGGTGGC	120
DB	64	GGTCCCGGGCAATTCCTCGATGTCACAGCGCTGCAATACGGCTTCTGGCGTGTAC	123
QY	121	GTCTATTGC--TGGGCTTCAGACCGGACTGGCGATTATTTGCTGGTTGGCG-TGGG	177
DB	124	TGGAACGCCCTGGGCTGAGCTGGGCTGATCATGCGATCGCATCGCCCGCG	183
QY	178	GTTGGACGCTATTTTCCCGCTCAGTATGCGTTGAAGTGTAAAGTGGCGAGCGCG	237
DB	184	GTCGGTCCGCTCCTGGCTGCCTCGGCCACGGGTTCCAGGTGATCAAGTGGTTGGCGTC	243
QY	238	GCTTACTTGTGCTGGGANTCCAGAGTGGCGCGGCT-----GGTGAATTGAC	291

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 14:07:39 ; Search time 1953.09 Seconds

(without alignments)
9403.168 Million cell updates/sec

Title: US-09-847-392-1_COPY_557_1171

Perfect score: 615

Sequence: 1 atgaccttggaatggtgtt.....tagcatggcgagcgatgcg 615

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estlin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estci:*

9: gb_estli:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estcom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	127	20.7	473	28	AQ859250 nbeb00111
C 2	85.4	13.9	1095	28	B2548406 pacs1-60
C 3	78.6	12.8	769	28	AQ858095 nbeb00111
C 4	55.8	9.1	1199	28	B2566001 pacs2-164

C 5	53	8.6	231	14	CB042356
C 6	41	6.7	849	13	EX462111
C 7	40.6	6.6	1039	29	CNS016GW
C 8	39.8	6.5	990	29	AY401573
C 9	39.4	6.4	712	13	EX416727
C 10	39.2	6.4	1101	29	CNS0021D
C 11	38.8	6.3	685	13	BY724027
C 12	38.6	6.3	642	9	AL628057
C 13	38	6.2	510	12	BG664436
C 14	37.6	6.1	475	28	BH209707
C 15	37.6	6.1	798	10	BE659545
C 16	37.2	6.0	885	13	EX425603
C 17	37.2	6.0	1159	29	CNS015XR
C 18	37	6.0	264	10	AW347581
C 19	37	6.0	308	10	AW266920
C 20	37	6.0	437	12	BG688576
C 21	37	6.0	450	10	AW482031
C 22	37	6.0	495	10	BF606721
C 23	37	6.0	516	10	BE663253
C 24	37	6.0	522	12	BM087647
C 25	37	6.0	539	12	BM253190
C 26	37	6.0	573	10	BE665319
C 27	37	6.0	573	12	BI773985
C 28	37	6.0	585	12	BI774021
C 29	37	6.0	599	14	CB466945
C 30	37	6.0	603	9	AV596505
C 31	37	6.0	629	14	CB462760
C 32	37	6.0	1201	13	EX444391
C 33	36.8	6.0	1201	13	EX385008
C 34	36.4	5.9	1661	12	BM473385
C 35	36	5.9	348	29	CE781017
C 36	36	5.9	453	13	BY272658
C 37	35.8	5.8	997	29	CNS006DN
C 38	35.6	5.8	474	9	AL518216
C 39	35.4	5.8	154	9	AV847594
C 40	35.4	5.8	439	10	BE346900
C 41	35.4	5.8	547	9	AI965398
C 42	35.4	5.8	668	14	CF179229
C 43	35.2	5.7	304	9	AL643781
C 44	35.2	5.7	1201	13	EX403316
C 45	35	5.7	931	13	BQ922252

ALIGNMENTS

RESULT 1
AQ859250/c

LOCUS

DEFINITION

AQ859250 473 bp DNA linear GSS 03-NOV-1999
nbeb001114r CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica
cultivar-group) genomic clone nbeb001114r, genomic survey
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 473)

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GSAAACAGCTATGACCATG

Class: BAC ends

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

High quality sequence start: 41
High quality sequence stop: 415.

FEATURES

source

1. .473

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbe001114f"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"

/clone_lib="CUGI Rice BAC Library (ECORI)"

/notes="vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,236 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Query Match 20.7%; Score 127; DB 28; Length 473;

Best Local Similarity 92.4%; Pred. No. 6.5e-27;

Matches 145; Conservative 0; Mismatches 10; Indels 2; Gaps 1;

Qy 1 ATGACCTTAGAATGGTGGTTGGCTACCTGCTGACATCGATCATTTAAGCGTGCACA 60

Db 229 ATGACCTTAGAATGGTGGTTGGCTACCTGCTGACATCGATCATTTAAGCGTGCACA 170

Qy 61 GGCTCTGGTCAATCAACATATGACCACTGCTCAACACGGTAT--CCGGCCGGTG 118

Db 169 GGCTCTGGTCAATCAACATATGACCACTGCTCAACACGGTATCGCGCGCGATA 110

Qy 119 GCGTCTATTGCTGGGCTTCAGACCGGACTGGCGATT 155

Db 109 AAGTCTATTGCTGGGCTTCAGACCGGACTGGCGATT 73

RESULT 2

B2548406

LOCUS

DEFINITION B2548406 1095 bp DNA linear GSS 17-DEC-2002

pacsl-60_1079.s1 pacsl-60 Pseudomonas aeruginosa genomic clone

pacsl-60_1079, genomic survey sequence.

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

Qy B2548406.1 GI:27151987

Db B2548406.1 GI:27151987

University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 20622216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES

source

1. 1095
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsl-60_1079"
/clone_lib="pacsl-60"
/note="clinical isolate 1-60 Whole genomic shotgun library."

ORIGIN

Query Match 13.9%; Score 85.4; DB 28; Length 1095;

Best Local Similarity 54.1%; Pred. No. 3.7e-14;

Matches 217; Conservative 0; Mismatches 181; Indels 3; Gaps 2;

Qy 218 TGTGAGTGGGCGGCGGCTTACTTGAATTTGGCTGGGAATCCAGCAGTGGC--GGCG 275

Db 41 TGATCAAGTGGTTCGGCTGGCTACCTGGTGTACCTGGCGTGGCGAGTGGCAGGCG 100

Qy 276 CGCTGTGCAATTGACCTTAAATCGCTGGCTCTACTCAATCG--CGTCGACATTTGTC 334

Db 101 CGCCACAGGCTTGGAGCCGATGCGAAGCGCTCTGGGGCGACCGTTGACCTGGTGC 160

Qy 335 AGCGGCGAGTTTGTGATCTCACCATCCAAAGTATTGTCTTTCTGGCGGGCTAT 394

Db 161 TCGTGTGTTTCTGTGTCACGCCAATCCCAAGGCGGTGATCTTCATGCTCGCGTGC 220

Qy 395 TTCCGCAATTTCATGCTGCGCAACAGCGCAATGATCAGTATATCGTCTCGGCGCTCA 454

Db 221 TGCGCGAGTTTCATGCGCGCGCGCGCGCTGCTGGCGCAATACCTGATCATGCGCGCA 280

Qy 455 CCATATTGCTGCTGATATTGATGATGATGATGATGATGATGATGATGATGATGATG 514

Db 281 CCATGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 340

Qy 515 CTCTATGGAATTAAAGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 574

Db 341 TAGCGTACTGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 615

Qy 575 TTATGCTGTGGAGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 615

Db 401 TCGTGTGCTGCG 441

RESULT 3

AQ858095

LOCUS

DEFINITION

AQ858095 769 bp DNA linear GSS 03-NOV-1999

nbe001114f CUGI Rice BAC Library (ECORI) Oryza sativa (japonica

cultivar-group) genomic clone nbe001114f, genomic survey

sequence.

Qy AQ858095

Db AQ858095

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

AQ858095 769 bp DNA linear GSS 03-NOV-1999

nbe001114f CUGI Rice BAC Library (ECORI) Oryza sativa (japonica

cultivar-group) genomic clone nbe001114f, genomic survey

sequence.

Qy AQ858095

Db AQ858095

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Qy AQ858095.1 GI:6208552

Db AQ858095.1 GI:6208552

Tel: 864 656 7288
Fax: 864 656 4293

Email: rwing@clmson.edu

Seq primer: TAATACGACTCATATAGG

Claes: BAC ends

High quality sequence start: 42

High quality sequence stop: 454.

FEATURES

Location/Qualifiers
1..769
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbeb0011114f"
/cisseu_type="leaf"
/lab_host="E. coli DH10B"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/note="Vector: pBACindigo; Site 1: EcoRI; Site 2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9%. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."

source

1..769
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="nbeb0011114f"
/cisseu_type="leaf"
/lab_host="E. coli DH10B"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/note="Vector: pBACindigo; Site 1: EcoRI; Site 2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9%. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."

ORIGIN

Query Match 12.8%; Score 78.6; DB 28; Length 769;
Best Local Similarity 94.5%; Pred. No. 3.4e-12;
Matches 103; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
QY 1 ATGACCTTAGAATGTTGGTTTCCCTACTGCTGATCGATCATCTTTTAAAGCTGTGCGCA 60
Db 658 ATGACCTTAGAATGTTGGTTTCCCTACTGCTGATCGATCATCTTTTAAAGCTGTGCGCA 715
QY 61 GCCTGTGTGCAATCAACACTATGACCACTCTGCTGACCACTGTTATC 109
Db 716 GCCTGTGTGCAATCAACACTATGACCACTCTGCTGACCACTGTTATC 764

RESULT 4

LOCUS BZ566001/c 1199 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_5858.xl pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_5858, genomic survey sequence.
ACCESSION BZ566001
VERSION BZ566001.1 GI:27194705
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1199)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) in press
COMMENT Contact: Chris K. Raymond

Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1..1199
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_5858"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN

Query Match 9.1%; Score 55.8; DB 28; Length 1199;
Best Local Similarity 50.1%; Pred. No. 3.3e-05;
Matches 195; Conservative 0; Mismatches 187; Indels 7; Gaps 2;
QY 128 GCTGGGTTTCAGACCGGACTGGCGATTATTTGCTGCTGGTGGGTTGGGACGC 187
Db 649 GCTGGCCTTCAGATCGGCTGCTGACAGATACCCATTGCTCTGCCGGGTCGGTACG 590
QY 188 TATTTTCC--CGCTCAGTATTGCTTGAAGTGTGAAGTGGGAGGCGCGCTTACTT 245
Db 589 TTAGTGGCAGCGTCGCGACTGCTTACAGCATGATCAAGTGTTCGGCGCTGACCTACAT 530
QY 246 GATTGGCTGGGAATCAGCAGTGGC-----GGCGCGCTGGTGCATTTGACCTTAAATCG 300
Db 529 GGTGTACTCTGTGTGGCGCAGTGGCAGCGCGCCACAGACCTTTGAGCAGCGATGGCG 470
QY 301 CTGGCCTTACTCAATCGGTCGACATTTGTTCCAGCGCGCAGTTTGTGATCTCACG 360
Db 469 ACCGGCCTCTGGTGCGACCGTTGACCCCTGGTGTGCTGCTGCTGACCGACAC 410
QY 361 AATCCCAAAAGTATTGTTTCTGGGGGGGCTATTTCCGCAATTCATCATGCGCAACAG 420
Db 409 AATCACAAGACGGTGTCTGATGCTCGGGTGGTCCCGAGTTCAGCGACCGCATCAT 350
QY 421 CCGCAATGATGACGATATATGCTGCTGGCGGTACCACTATTGTGTCGATATTATG 480
Db 349 CCGTGTGTGCGAAGACCTGATCATGCGCGCGCCCATGATCGTCTGTCGTGATCGTC 290
QY 481 ATGATCGGTTACGCCACCCCTTGTCTCAACG 509
Db 289 ATGTCCAGTACACCGGGCTGGCTGCGCG 261

RESULT 5

LOCUS CB042356/c 231 bp mRNA linear EST 15-JAN-2003
DEFINITION 4007198 BARC-EMERAPA 316BOV Bos indicus cDNA clone 316BOV_1P12
Unknown, mRNA sequence.
ACCESSION CB042356
VERSION CB042356.1 GI:27761601
KEYWORDS EST.
SOURCE Bos indicus (zebu)
ORGANISM Bos indicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 231)
AUTHORS da Mota, A.F., Sonstegard, T.S., Van Tassel, C.P., Matukumalli, L.K.,
Wood, D.L., Capuco, A.V., Brito, M.A.P., Martinez, M.L., Connor, E.E.,
Machado, M.A. and Coutinho, L.L.
TITLE Construction and Characterization of cDNA Libraries Generated from
Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
indicus) Cattle
Unpublished (2002)
JOURNAL

COMMENT

Contact: Adilson F. da Vota
Gene Evaluation and Mapping Laboratory
USDA, ARS, Animal and Natural Resources Institute
Bldg. 200 Rm3 BAC-East, Beltsville, MD 20705, USA
Tel: 3015048456
Fax: 3015048414
Email: amote@cpnpl.embrapa.br
Single pass sequencing. Bases called and trimmed with phred
0.00925 using options: trim-alt, trim-fasta. Vector identified
by cross_match using options -minmatch 12 -minscore 18
PCR Primers
FORWARD: GTTTCCAGTCACGACGTTG
BACKWARD: TGAGCGGATAACAATTTCACACG
Plate: 1 row: P column: 12
Seq primer: GTTTCCAGTCACGACGTTG
High quality sequence stop: 231.
Location/Qualifiers
1. 231

FEATURES

source

/organism="Bos indicus"
/mol_type="mRNA"
/strain="Brazilian Dairy Gir"
/db_xref="taxon:9915"
/clone="316BOV_1P12"
/sex="female"
/tissue_type="mammary"
/cell_type="epithelium"
/dev_stage="involved"
/lab_host="K-12"
/clone_lib="BAC-EMBRAPA 316BOV"
/notes="Organ: mammary; Vector: pUC19; Site 1: SmaI;
Site 2: SmaI; this mammary-derived cDNA library was
created as part of a collaborative project between the ARS
Gene Evaluation and Mapping Laboratory and the EMBRAPA
Dairy Cattle Research Center under the sponsorship of
USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
extracted on 6/8/02, RT with Superscript II at 37 deg C,
annealing temperature, PCR with 15mer AMS."

ORIGIN

Query Match 8.6%; Score 53; DB 14; Length 231;
Best Local Similarity 56.6%; Pred. No. 8.7e-05;
Matches 98; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
QY 322 CGACATTGTTCCAGCGCAGTCTTGTGATCTCACCATCCCAAGATTGTTT 381
Db 201 CGCGTGGTTCCTGTTGGATTGTCGACCAATCTGACGACCCGAGGCGATCCTGTC 142
QY 382 CTGCGCGCGTATTTCGCAATTATCATGCGCGCAACGCGCAACTGATGCAATATC 441
Db 141 ATGGTGGCGTGTGCGCGAGTTCATGATCGCGCAAGCGCTGTTTCCAACTCGGC 82
QY 442 GTCTCGGGGTACCACTATTGGTGGATATATGATGATCGGTACGC 494
Db 81 ATCTGGCGGCGAGATGGTGTGTCGACCTGATGTCGTCAGCGGTATGC 29

RESULT 6

BX462111/c

LOCUS 849 bp mRNA linear EST 22-MAY-2003
DEFINITION BX462111 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
clone CS0DG001P07 5-PRIME, mRNA sequence.
ACCESSION BX462111
VERSION BX462111.1 GI:31029391
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 849)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9966.f For
more information about this cluster, see

http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DG001CH04QP1&cluster=9966.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1500
Faraday Avenue Genoscope sequence ID : CS0DG001CH04QP1.

FEATURES

source

Location/Qualifiers
1. 849
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DG001P07"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN

Query Match 6.7%; Score 41; DB 13; Length 849;
Best Local Similarity 21.3%; Pred. No. 0.77;
Matches 53; Conservative 86; Mismatches 110; Indels 0; Gaps 0;
QY 105 TTATCCGCGCGTGGCTATTCTGGGTTCCAGACCGACTGGCGATTTCATTGTGTC 154
Db 500 TTGCTGCTTTTGCCTGCTTTGCTTTAGCCTTTGCGCTATTGACGGCGCGCATTTGC 441
QY 165 TGTGTGGCTGGGTTGGAGCTATTTCCTCGCTCAGTGTGCTTTTGAAGTTGAA 224
Db 440 TTGCTGCKGCKGCKGCKGCKGCKGCKGCKGCKGCKGCKGCKGCKGCKGCKGCKG 381
QY 225 GTGGCAGCGCGCGGTACTTATTGTTGGCTGGGAATCCAGCAGTGGCGCGCGTGTGC 284
Db 380 KKSSTASGSGTSGSKKAKKSKKSKKSAKSGSKGSKGSKGSKGSKGSKGSKGSKG 321
QY 285 AATTACCTTAATCGTGGCTCTACTCATCGCTCGACATTTGTCAGCGCGCAGT 344
Db 320 TGSNCNKKSAATATGSGSKKSKKSKTASTTSSKSKSKKSKSKSKSKSKSKSK 261
QY 345 TTTTGTGAA 353
Db 260 CKKGSKKKA 252

RESULT 7

CNS016GW

LOCUS

DEFINITION CNS016GW 1039 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence T7 end of BAC
BACN16011 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION

AL106730

VERSION

AL106730.1 GI:5623541

KEYWORDS

GSS.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1039)

Genoscope.

REFERENCE

Direct Submission

AUTHORS

TITLE

JOURNAL

COMMENT

Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES source

1..685
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="A230077K02"
 /sex="male"
 /tissue_type="hypothalamus"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, adult male hypothalamus"
 /note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTNN 3'], cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATTAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 6.3%; Score 38.8; DB 13; Length 685;
 Best Local Similarity 53.0%; Pred. No. 3.1;
 Matches 105; Conservative 0; Mismatches 92; Indels 1; Gaps 1;

QY 24 CTACCTGCTGACATGATCTTTTACGCTGCGCCAGCTCTGTGCAATCAACACTAT 83
 Db 443 CTCAATTGTAGCGTGCACCACTGACACACTTTCTCTAGCCCTCTACTTC-CTGCGCGTCAT 501

QY 84 GACCACCTCGCTCAACCAACGGTTATCCGCGCGGTGGCGTCTATTGCTGGGGTTCAGACCG 143
 Db 502 GAGCGCGGACCGATACCTGTGTGTTCTGCGCCACACAGAGTCGCGCGGGTGTCCGGCGG 561

QY 144 GACTGGCGATTATATGCTGTGGTGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGT 203
 Db 562 CACTTACGGTGCAGCGCGTGTGTCAGTCTGCGGTGTGGCGGTGGCGGTGGCGGTGGCGGT 621

QY 204 GATTGGTGTGAAGTGT 221
 Db 622 GCTGCCCTTTCGGTATT 639

RESULT 12 AL628057/c

LOCUS
 DEFINITION AL628057 XGC-gastrula silurana tropicalis cDNA clone TGas022011 5', mRNA sequence.
 ACCESSION AL628057
 VERSION AL628057.2 GI:38448162
 KEYWORDS EST.
 SOURCE Silurana tropicalis (western clawed frog)
 ORGANISM Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Silurana.
 REFERENCE 1 (bases 1 to 642)
 AUTHORS Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.
 TITLE Sanger Xenopus tropicalis EST project 2001 (11_2003)
 JOURNAL Unpublished (2003)
 COMMENT On Nov 2, 2001 this sequence version replaced gi:16597540.
 Contact: Huckle E
 Sanger Institute
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trp@sanger.ac.uk
 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.
 cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.
 Vector: pCS107; Site 1: EcoRI; Site 2: NotI
 Host: Escherichia coli XL1-blue
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS_SEQUENCE ID: TGas022011.p1kSP6
 Sequencing primer: SP6.
 Location/Qualifiers
 1..642
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TGas022011"
 /dev_stage="gastrula (stages 10.5-12 mixed)"
 /lab_host="Escherichia coli XL1-blue"
 /clone_lib="XGC-gastrula"
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

FEATURES

source

1..642
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TGas022011"
 /dev_stage="gastrula (stages 10.5-12 mixed)"
 /lab_host="Escherichia coli XL1-blue"
 /clone_lib="XGC-gastrula"
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 6.3%; Score 38.6; DB 9; Length 642;
 Best Local Similarity 50.8%; Pred. No. 3.5;
 Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 295 AAATCGCTGGCTCTACTCAATCGCTCGACATTTGTCACGCGCGAGTTTGTGAAT 354
 Db 252 AAATAGCATGAATTTGTTCCATTGCTTCAAGTTTGGTTTCATGGCTCAGTTTGGCTCTGG 193

QY 355 CTCACCAATCCAAAGATTGTTGTTCTGGCGCGCTATTTCGCAATTCATCATGCCG 414
 Db 192 TTCCTTCGGCACTACTGATGATGATTCAGGACACGATACCTGATCCTCCCTCTCT 133

QY 415 CAACAGCGCACTGATGACGATATATCGTCTGGCGTCCACCTATTGTGTCGATATT 474
 Db 132 CAGGAGCAGAGATGGCAGCGAGTAGCGGGAGACCCGCTGCATTTGCTGGATGGAGACT 73

QY 475 A 475
 Db 72 A 72

RESULT 13

LOCUS
 DEFINITION BG664436
 510 bp mRNA linear EST 30-APR-2001
 DRABDH02 Rat DRG Library Rattus norvegicus cDNA clone DRABDH02 5',

mRNA sequence.
 ACCESSION BC664436
 VERSION BC664436.1 GI:13886358
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
 1 (bases 1 to 510)
 Xiao, H.S., Huang, Q.H., Zhang, F.X., Bao, L., Lu, Y.J., Guo, C., Yang, L., Huang, W.J., Fu, G., Xu, S.H., Cheng, X.P., Yan, Q., Zhu, Z.D., Zhang, X., Chen, Z., Han, Z.G. and Zhang, X.
 Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
 Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
 22056133
 CONTACT: Zhang Xu
 Laboratory of Sensory System
 Institute of Neuroscience
 320 Yue Yang Road, Shanghai 200031, P.R.China
 Tel: 86-21-64748700-121
 Fax: 86-21-64713446
 Email: xu.zhang@ion.ac.cn
 This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@sh.cn)
 PCR Primers
 FORWARD: T3
 BACKWARD: T7
 Seq primer: T3
 POLYA=No.
 FEATURES
 source
 Location/Qualifiers
 1..510
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="DRABDH02"
 /sex="male"
 /tissue_type="dorsal root ganglion"
 /dev_stage="adult"
 /clone_lib="Rat DRG Library"
 ORIGIN
 Query Match 6.2%; Score 38; DB 12; Length 510;
 Best Local Similarity 53.3%; Pred. No. 4.6;
 Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
 QY 378 GTTCTGGCGGCGCTATTCCGCAATTCATGATCCGCAACAGCGCACTGATGCACTA 437
 Db 204 GTTATTCCAGCGGAGATCTCCCAACATGATGACCAATGCGACACCACTTAATGA 263
 QY 438 TATCGTCTGGCGGCTCACCACATTTGCTGCATATTATTTGATGATCGTTACGCCAC 497
 Db 264 CCTTGTTATCAGGATGGCTCTTCTCGAGAGATCTTGTGAGAGCAATCTGGATCC 323
 QY 498 CTTGCTTCAACGATGCTCTATGATTA 527
 Db 324 CAATGCTAAGGAGATTGTTCTCTGGGGTGAA 353
 RESULT 14
 BH209707/c
 LOCUS BH209707 475 bp DNA linear GSS 15-NOV-2002
 DEFINITION Sml-41P23.TF Sml Schistosoma mansoni genomic clone Sml-41P23, genomic survey sequence.
 ACCESSION BH209707
 VERSION BH209707.1 GI:16388592
 KEYWORDS GSS.
 SOURCE Schistosoma mansoni
 ORGANISM Schistosoma mansoni
 Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
 1 (bases 1 to 475)
 Shetty, J., Simpson, A., Malek, J., Koo, H., LoVerde, P.T. and El-Sayed, N.M.
 Use of end sequences from Schistosoma mansoni (Puerto Rico strain) Sml BAC library for gene discovery and map construction
 Unpublished (2001)
 Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: nelsayed@tigr.org
 Clones are derived from the Schistosoma mansoni (Puerto Rico) Sml BAC library. For clone availability, please contact Dr. Najib El-Sayed at TIGR (nelsayed@tigr.org) or Dr. Phillip LoVerde at State University of New York, Buffalo, New York, USA (loverde@buffalo.edu)
 Seq primer: M13 For
 Class: BAC ends.
 FEATURES
 Location/Qualifiers
 1..475
 /organism="Schistosoma mansoni"
 /mol_type="genomic DNA"
 /strain="Puerto Rico"
 /db_xref="taxon:6183"
 /clone="Sml-41P23"
 /clone_lib="Sml"
 /note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed in the laboratory of Dr. Denis Le Paslier at the Fondation Jean Dausset, CCRP, Paris, France. Briefly, Schistosoma mansoni agarose embedded DNA was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 7.95 X the haploid genome. Further information can be found in Le Paslier et al. (2000) Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library. Genomics 65: 87-94."
 ORIGIN
 Query Match 6.1%; Score 37.6; DB 28; Length 475;
 Best Local Similarity 48.2%; Pred. No. 5.8;
 Matches 106; Conservative 0; Mismatches 114; Indels 0; Gaps 0;
 QY 377 TGTTCCTGGCGGCGCTATTCCGCAATTCATGCGCAACAGCGCACTGATCAGT 436
 Db 352 TGATCCTGCTGTTCTGCTGATGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTG 293
 QY 437 ATATCGTCTCGGCGTCACCACATTTGCTGCGATATTATTGATGATCGGTTACGCCA 496
 Db 292 TGCTGCTGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGA 233
 QY 497 CCCTTGCTCAACGGATGCTCTATGGATTAAAGGACCAAGCAGATGAAGCGCTGAATA 556
 Db 232 TGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATA 173
 QY 557 AGATTTTCGGCTCGTTGTTTATGCTGCTGGAGCGCTGTT 596
 Db 172 AGATTTGGACATGATGTTGTTTGTGACATGACGATGAT 133
 RESULT 15
 BE559545/c
 LOCUS BE559545 798 bp mRNA linear EST 24-MAY-2001
 DEFINITION GW700010A1.0A1 Gm-r1070 Glycine max cDNA clone Gm-r1070-3841 3', mRNA sequence.
 ACCESSION BE559545
 VERSION BE559545.1 GI:9985437
 KEYWORDS EST.

SOURCE
ORGANISM

Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 798)
Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
Erpelding,J., Rapp,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
JOURNAL
COMMENT
Other_ESTs: AI965398 corresponding to Gm-cl016-1844 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES
source

Location/Qualifiers
1..798
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-3841"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from that
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones from
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Retzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/soybean/index.html
Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Keck Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."

ORIGIN

Query Match 6.1%; Score 37.6; DB 10; Length 798;
Best Local Similarity 50.4%; Pred No. 7,9;
Matches 67; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
398 CGCAATTATCATGCGCGCAACGCGCACTGATGCGTCTCGCGTCAACCA 457
Db
738 CNNNATNNNNCGTAAACACCCNNNNCTNNNTGTTATTATTCCTCACTATTATCA 679
QY 458 CTATTGTGCTGATATTGATGATGCGGTAGCGCACCCCTTGCTCAACGGATTGCTC 517
Db 678 CTATCGTTATCGTTATCGTTGTTATCACTCTTCATCCCGTTCCCAACACAGAGTCT 619

QY 518 TATGATTAAAGG 530
Db 618 TATGTTCTCGAAG 606

Search completed: March 28, 2004, 18:21:30
Job time : 1968.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:08:16 ; Search time 60 Seconds

(without alignments)
965.371 Million cell updates/sec

Title: US-09-847-392-2

Perfect score: 1061

Sequence: 1 MTEWFWAYLTSITLTP.....KIFGSLFMLVGALLASARHA 205

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1061	100.0	205	3	AAy79298 E. coli R
2	1061	100.0	205	3	AAy9597 E. coli L
3	830.5	78.3	206	6	ABU32233 Protein e
4	827.5	78.0	206	6	ABU28287 Protein e
5	818.5	77.1	206	6	ABU47858 Protein e
6	803.5	75.7	190	4	ABG29208 Novel hum
7	791	74.6	205	6	ABU44939 Protein e
8	698.5	65.8	206	6	ABU50100 Protein e
9	686	64.7	138	4	AAg98975 E. coli g
10	586	64.7	138	6	ABU14756 Protein e
11	459.5	43.3	205	6	ABU49040 Protein e
12	410	38.6	245	4	ABG25225 Novel hum
13	389	36.7	209	6	ABU38890 Protein e
14	368	34.7	220	6	ADA33616 Acinetoba
15	367.5	34.6	210	6	ABU41414 Protein e
16	354.5	33.4	211	6	ABU39959 Protein e
17	312.5	23.5	209	6	ABU23178 Protein e
18	291	27.4	181	6	ABU16949 Protein e
19	290.5	27.4	211	6	ABU19902 Protein e
20	260.5	24.6	171	4	ABG29211 Novel hum
21	227	21.0	346	4	ABG29210 Novel hum
22	223	21.0	210	6	ADA35521 Acinetoba
23	215.5	20.3	241	6	ADA35872 Acinetoba
24	198	18.7	208	6	ADA33824 Acinetoba
25	195.5	18.4	212	3	ABU1787 Escherich

25	170	16.0	240	6	ADA36250	Ada36250 Acinetoba
27	168.5	15.9	235	6	ADA36962	Ada36962 Acinetoba
28	156.5	14.8	201	6	ABM68999	Abm68999 Photorhab
29	155	14.6	153	2	AAr87527	Ar87527 Mel-linke
30	155	14.6	153	2	AAW73358	Aw73358 S. Colwel
31	150	14.1	207	4	AAG92664	Ag92664 C glutami
32	147	13.9	249	6	ADA34526	Ada34526 Acinetoba
33	146.5	13.8	205	6	ABM67507	Bm67507 Photorhab
34	144	13.6	130	3	ABM40407	Abm40407 Human ORF
35	144	13.6	130	5	ABP00692	Abp00692 Human ORF
36	134.5	12.7	226	4	AAG89911	Ag89911 C glutami
37	133.5	12.6	210	4	AAG81807	Ag81807 S. epider
38	133.5	12.6	224	5	ABP39763	Abp39763 Staphyloc
39	133	12.5	222	6	ADA35184	Ada35184 Acinetoba
40	126.5	11.9	213	6	ABM67782	Bm67782 Photorhab
41	126.5	11.9	223	3	ABM1786	Abm1786 Escherich
42	126	11.9	195	3	ABM1788	Abm1788 Escherich
43	124.5	11.7	214	6	ADA35408	Ada35408 Acinetoba
44	123.5	11.6	191	4	ABG29209	Abg29209 Novel hum
45	123.5	11.6	229	3	ABM11637	Abm11637 A. vitis

ALIGNMENTS

RESULT 1
AAy79298
ID AAY79298 standard; protein; 205 AA.
XX
AC AAY79298;
XX
DT 18-JUL-2000 (first entry)
XX
DE E. coli RhtB protein imparting homoserine resistance.
XX
KW Homoserine resistance; RhtB protein; L-homoserine; L-alanine;
KW L-isoleucine; L-valine; L-threonine.
XX
OS Escherichia coli.
XX
PN EP994190-A2.
XX
PD 19-APR-2000.
XX
PF 20-SEP-1999; 99EP-00118581.
XX
PR 13-OCT-1998; 98RU-00118425.
XX
PA (AJIN) AJINOMOTO CO INC.
XX
PI Livshits VA, Zakataeva NP, Aleoshin VV, Belareova AV;
PI Tokhmakova IL;
DR WPI; 2000-273530/24.
XX
N-PSDB; AAZ94405.
PT Novel RhtB protein, useful for generation of L-homoserine resistance in
PT Escherichia bacteria and large-scale production of e.g. L-homoserine and
PT L-alanine.
XX
Claim 1; Page 11-12; 14pp; English.

The present sequence is that of the novel Escherichia coli X-12 protein, RhtB, which participates in resistance to L-homoserine. Amplification of the rhtB gene (see AAY94405) results in an improvement of the amino acid productivity of E. coli. The invention provides: a protein comprising the present sequence or including a deletion, substitution, insertion and/or addition of 1 or more amino acids and having the activity of making a bacterium resistant to L-homoserine; DNA encoding RhtB; a bacterium, especially of the genus Escherichia, in which L-homoserine resistance is enhanced by amplifying the copy number or increasing the expression rate of the rhtB DNA; the DNA being carried on a multicopy vector or on a transposon; and a method for producing an amino acid by cultivating the

CC bacterium in a culture medium to produce and accumulate the amino acid in
CC the medium from which it is recovered. The method is used for the
CC production of L-homoserine, L-alanine, L-isoleucine, L-valine or L-
CC threonine (all claimed)

XX
SQ Sequence, 205 AA;
Query Match 100.0%; Score 1061; DB 3; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.2e-112;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWRGV 60
DB 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWRGV 60
QY 61 GTLFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQSRHILFORAVFNLT 120
DB 61 GTLFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQSRHILFORAVFNLT 120
QY 121 NPKSIVFLAALFPQFTMPQPOLMQYIVLGVTTIIVVDIIVMIGVATLAQRIALWIKPKQ 180
DB 121 NPKSIVFLAALFPQFTMPQPOLMQYIVLGVTTIIVVDIIVMIGVATLAQRIALWIKPKQ 180
QY 181 MKALNKIFGSLFVLGALLASARHA 205
DB 181 MKALNKIFGSLFVLGALLASARHA 205

RESULT 2
AAV99597
ID AAV99597 standard; protein; 205 AA.

XX
AC AAV99597;
XX
DT 08-SEP-2000 (first entry)
XX
DE E. coli L-homoserine resistance protein, RhtB.
XX
DE L-homoserine resistance; L-homoserine synthesis; rhtB; L-threonine;
KW L-valine; L-leucine.
XX
OS Escherichia coli.
XX
PN EPI013765-A1.
XX
PD 28-JUN-2000.
XX
PF 20-DEC-1999; 99EP-00125406.
XX
PR 23-DEC-1998; 98RU-00123511.
XX
PA (AJIN) AJINOMOTO KK.

PI Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;
XX
XX WPI; 2000-414602/36.
DR
DR N-PSDB; AAA48442.

XX
PT Novel Escherichia bacterium having enhanced L-threonine resistance due to
PT enhanced RhtC protein activity, used to produce L-threonine, L-
PT homoserine, L-valine and L-leucine.

XX
PS Claim 2; Page 13-14; 24pp; English.

XX
XX The present sequence is the L-homoserine resistance protein, RhtB, from
XX Escherichia coli. The coding sequence may be used to impart L-homoserine
XX resistance on E. coli bacteria, which would be useful for producing a
XX high yield of L-homoserine. L-homoserine resistance means that the
XX bacteria will be able to grow on a minimal medium containing L-homoserine
XX at a concentration at which the corresponding wild-type strain would not
XX grow. Since the transformed bacteria can grow on the minimal medium, it
XX can synthesize L-homoserine, which accumulates. The accumulated amino
XX acids can then be removed from the culture medium. The bacterium of the

CC present invention may also be used to synthesize L-threonine, L-valine
CC and L-leucine at increased levels
XX
SQ Sequence 205 AA;

Query Match 100.0%; Score 1061; DB 3; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.2e-112;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWRGV 60
DB 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWRGV 60
QY 61 GTLFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQSRHILFORAVFNLT 120
DB 61 GTLFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQSRHILFORAVFNLT 120
QY 121 NPKSIVFLAALFPQFTMPQPOLMQYIVLGVTTIIVVDIIVMIGVATLAQRIALWIKPKQ 180
DB 121 NPKSIVFLAALFPQFTMPQPOLMQYIVLGVTTIIVVDIIVMIGVATLAQRIALWIKPKQ 180
QY 181 MKALNKIFGSLFVLGALLASARHA 205
DB 181 MKALNKIFGSLFVLGALLASARHA 205

RESULT 3
ABU32233
ID ABU32233 standard; protein; 206 AA.

XX
AC ABU32233;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #17760.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Klebsiella pneumoniae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.

XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.

XX
PA (ELIT-) ELITRA PHARM INC.

XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX
XX WPI; 2003-029926/02.
DR
DR N-PSDB; ACA36103.

XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

XX
PS Claim 25; SEQ ID NO 60157; 1766pp; English.

XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the

antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 206 AA;

Query Match 78.3%; Score 830.5; DB 6; Length 206;
Best Local Similarity 81.2%; Pred. No. 2.6e-86;
Matches 169; Conservative 13; Mismatches 21; Indels 5; Gaps 3;

QY 1 MTEWFAVLLTSIIILSPGSGAINMTTSLNHGYPAGGVYCWASDRG---DSYCAGW 57
DB 1 MTEWFAVLLTSIIILSPGSGAINMTTSLNHGYPAGGVYCWASDRG---DSYCAGW 58
QY 58 RGVGLTFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQRRHLFORAVFV 117
DB 59 VGLGLTFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQRRHLFORAVFV 118
QY 118 NLTPKSVIFLAALFPQFIPOQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW 177
DB 119 NLTPKSVIFLAALFPQFIPOQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW 178
QY 178 PKQKALNKVFGSLFMLVGALLASARHA 205
DB 179 PKQKALNKVFGSLFMLVGALLASARHA 206

RESULT 4
ABU28287
ID ABU28287 standard; protein; 206 AA.
AC ABU28287;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #13814.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Enterobacter cloacae.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.
PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA32157.
XX New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 56211; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 206 AA;

Query Match 78.0%; Score 827.5; DB 6; Length 206;
Best Local Similarity 79.1%; Pred. No. 5.7e-86;
Matches 167; Conservative 15; Mismatches 18; Indels 11; Gaps 3;

QY 1 MTEWFAVLLTSIIILSPGSGAINMTTSLNHGYPAGGVYCWASDRGSDSYCAGWR-- 58
DB 1 MTEWFAVLLTSIIILSPGSGAINMTTSLNHGYPAGGVYCWASDRGSDSYCAGWR-- 55
QY 59 ---GVGTLFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQRRHLFORA 114
DB 56 LVGIGLTLFSRSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQRRHLFORA 115
QY 115 VFVNLTPKSVIFLAALFPQFIPOQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW 174
DB 116 VFVNLTPKSVIFLAALFPQFIPOQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW 175
QY 175 IKGPQKALNKVFGSLFMLVGALLASARHA 205
DB 176 IKGPQKALNKVFGSLFMLVGALLASARHA 206

RESULT 5
ABU47858
ID ABU47858 standard; protein; 206 AA.

XX AC ABU47858;
 XX DT 19-JUN-2003 (first entry)
 XX DE Protein encoded by Prokaryotic essential gene #33385.
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX OS Salmomella typhi.
 XX PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 25-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX DR WPI; 2003-029926/02.
 XX DR N-PSDB; ACA51728.
 XX PT New antisense nucleic acids, useful for identifying proteins or screening
 XX PT for homologous nucleic acids required for cellular proliferation to
 XX PT isolate candidate molecules for rational drug discovery programs.
 XX PS Claim 25; SEQ ID NO 75782; 176pp; English.
 XX CC The invention relates to an isolated nucleic acid comprising any one of
 XX CC the 6213 antisense sequences given in the specification where expression
 XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
 XX CC encoding a polypeptide whose expression is inhibited by the antisense
 XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 XX CC polypeptide or its fragment whose expression is inhibited by the
 XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
 XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 XX CC proliferation or the activity of a gene in an operon required for
 XX CC proliferation; (7) identifying a compound that influences the activity of
 XX CC the gene product or that has an activity against a biological pathway
 XX CC required for proliferation, or that inhibits cellular proliferation; (8)
 XX CC identifying a gene required for cellular proliferation or the biological
 XX CC pathway in which a proliferation-required gene or its gene product lies
 XX CC or a gene on which the test compound that inhibits proliferation of an
 XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 XX CC compound's activity; (11) a culture comprising strains in which the gene
 XX CC product is overexpressed or underexpressed; (12) determining the extent
 XX CC to which each of the strains is present in a culture or collection of
 XX CC strains; or (13) identifying the target of a compound that inhibits the
 XX CC proliferation of an organism. The antisense nucleic acids are useful for
 XX CC identifying proteins or screening for homologous nucleic acids required
 XX CC for cellular proliferation to isolate candidate molecules for rational
 XX CC drug discovery programs, or for screening homologous nucleic acids
 XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 XX CC *K. pneumoniae* or *F. aeruginosa*. The present sequence is encoded by one of
 XX CC the target prokaryotic essential genes. Note: The sequence data for this
 XX CC patent did not form part of the printed specification, but was obtained
 XX CC in electronic format directly from WIPO at
 XX CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 206 AA;
 XX Query Match 77.1%; Score 818.5; DB 6; Length 206;
 XX Best Local Similarity 78.7%; Pred. No. 6.1e-85;

Matches 166; Conservative 13; Mismatches 21; Indels 11; Gaps 3;
 QY 1 MTLWMFAYLLTSIILTSFGSGAINTMTTSLNHGYPAGVYCWASDRGDSYCAGWR-- 58
 DB 1 MTFEWMFAYLLTSTLLSLSPGSGAINTMTTSLNHGY-RGAAASIAQLQTG----LGIHIV 55
 QY 59 ----GYCTLESRSVIAFEVLKWAAGAYLIWLGTQQWRAAGADLKLSTQSRRLHFORA 114
 DB 56 LVGVGLTLESRSLIAFEILKWAAGAYLIWLGTQQWRAAGADLHDLTLAGTQGRGLFKEA 115
 QY 115 VFVNLTPKSIIVFLAALFPQFIMPOQPQLMQYIVLGVTTIVVDIIVMIGYATLAORIAIW 174
 DB 116 IFVNLTPKSIIVFLAALFPQFIMPOQPQLAQYLLGVTTIVVDIIVMIGYATLAORIAIW 175
 QY 175 IKGPKQKALKNKIFGSLFVLVGLLALASARHA 205
 DB 176 IKGPKQKALKNKAFGSLFVLVGLLALASARHA 206
 RESULT 6
 ID ABG29208 standard; protein; 190 AA.
 XX AC ABG29208;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #29199.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 XX PR 31-MAR-2000; 2000US-00540217.
 XX PR 23-AUG-2000; 2000US-00649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS93395.
 XX CC New isolated polynucleotide and encoded polypeptides, useful in
 XX CC diagnostics, forensics, gene mapping, identification of mutations
 XX CC responsible for genetic disorders or other traits and to assess
 XX CC biodiversity.
 XX PS Claim 20; SEQ ID NO 59567; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
 XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 XX CC and in recombinant production of (II). The polynucleotides are also used
 XX CC in diagnostic as expressed sequence tags for identifying expressed
 XX CC genes. (I) is useful in gene therapy techniques to restore normal
 XX CC activity of (II) or to treat disease states involving (II). (II) is
 XX CC useful for generating antibodies against it, detecting or quantitating a
 XX CC polypeptide in tissue, as molecular weight markers and as a food
 XX CC supplement. (II) and its binding partners are useful in medical imaging
 XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
 XX CC involving aberrant protein expression or biological activity. The
 XX CC polypeptide and polynucleotide sequences have applications in
 XX CC diagnostics, forensics, gene mapping, identification of mutations
 XX CC responsible for genetic disorders or other traits to assess biodiversity
 XX CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 190 AA;

Query Match 75.7%; Score 803.5; DB 4; Length 190;
 Best Local Similarity 88.9%; Pred. NO. 2.8e-83;
 Matches 169; Conservative 3; Mismatches 13; Indels 5; Gaps 3;

QY 19 SPGSGAINTMTSLNHGYPAGVYCVMSDRGTG--DSYACAGWGVGTLSRSRVIAPEVLK 75
 DB 3 SPGSGAINTMTSLNHGY-RGAVASISAGLQTGLAIHVLVG-VGLGTLRSRVIAPEVLK 60
 QY 76 WAGAAYLWLGICQWRAAGADLKSLASTQSRHLFORAVFVNLTPKSIIVFLAALFPQF 135
 DB 61 WAGAAYLWLGICQWRAAGADLKSLASTQSRHLFORAVFVNLTPKSIIVFLAALFPQF 120
 QY 136 IMPOQPQLMQIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQKALNKIFGSLFMLV 195
 DB 121 IMPOQPQLMQIVLGVTTIVVDIIVMIGYATLAQRIALWIKGPKQKALNKIFGSLFMLV 180
 QY 196 GALLASARHA 205
 DB 181 GALLASARHA 190

RESULT 7
 ABU44939
 ID ABU44939 standard; protein; 205 AA.
 XX
 AC ABU44939;

XX 19-JUN-2003 (first entry)
 DT Protein encoded by Prokaryotic essential gene #30466.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Salmonella paratyphi.

XX WO200277183-A2.
 XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.
 XX 25-OCT-2001; 2001US-0342923P.
 XX 08-FEB-2002; 2002US-00072851.
 XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 XX N-PSDB; ACA48809.

XX New antisense nucleic acids, useful for identifying proteins or screening
 XX for homologous nucleic acids required for cellular proliferation to
 XX isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 72863; 1766bp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of
 XX the 6213 antisense sequences given in the specification where expression
 XX of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 205 AA;

Query Match 74.6%; Score 791; DB 6; Length 205;
 Best Local Similarity 76.8%; Pred. NO. 8.4e-82;
 Matches 162; Conservative 14; Mismatches 23; Indels 12; Gaps 4;

QY 1 MTEWFFAYLLTSIILTLSPGSAINTMTSLNHGYPAGVYCVMSDRGTGSDYCAQWR-- 58
 DB 1 MTFEWFAYLLTSLLSLSPGSAINTMTSLNHGY-RGAASISAGLQTG----LGIHV 55
 QY 59 ----GVGTLSRSRVIAPEVLKWAGAAAYLIWLGIQWRAAGADLKSLASTQSRHLFORA 114
 DB 56 LVGVGLGTLFSRSLLAFELIKWAGAAAYLIWLGIQWWR-AGADLHTLAQTQSRGLFKRA 114
 QY 115 VFVNLTPKSIIVFLAALFPQFIMPOQPQLMQIVLGVTTIVVDIIVMIGYATLAQRIALW 174
 DB 115 IFVNLTPKSIIVFLAALFPQFIMPOQPQLAQYLILGVTTIVVDIIVMIGYATLAXRTAAW 174
 QY 175 IKGPQKQKALNKIFGSLFMLVGLLASARHA 205
 DB 175 IKGPQKQKALNKAFGSLFMLVGLLASARHA 205

RESULT 8
 ABUS0100
 ID ABUS0100 standard; protein; 206 AA.

XX AC ABUS0100;
 XX 19-JUN-2003 (first entry)
 DT Protein encoded by Prokaryotic essential gene #35627.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Yersinia pestis.
 XX WO200277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.
08-FEB-2002; 2002US-00072851.
06-MAR-2002; 2002US-032699P.
(ELIT-) ELITRA PHARM INC.
Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
WPI: 2003-029926/02.
N-PSDB; ACA53970.
New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.
Claim 25; SEQ ID NO 78024; 1766pp; English.
The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits the
proliferation of an organism. The antisense nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
the target prokaryotic essential genes. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

Sequence 206 AA;
Query Match 65.8%; Score 598.5; DB 6; Length 206;
Best Local Similarity 63.3%; Pred. No. 3.1e-71;
Matches 136; Conservative 33; Mismatches 23; Indels 23; Gaps 3;
1 MTEWFWAYLLTSLTILSPGSGAINTWTSLNHHGYPAGGVYCWASDRGTGDSYCAGWR-- 58
1 MTLDWMLYLLTLLSLSPGSGAINTWTSLNHHGYPAGGVYCWASDRGTGDSYCAGWR-- 58
59 -----GVGTLFBSRVTAPEVLKWAAGAYLWLGICQWRAAGADLKSLASTQSR 108
50 LAHVILVGVGLGALVQSGLAFELKWLGAAYLWLGICQWRAAGADLKSLASTQSR 109
109 HLFQRAVFNLTNPKSIVFLAALFPQIPQPMQOYIVLGVTTIVDVIIVMIGVATLA 168
110 KLFKEAVFNLTNPKSIVFLAALFPQIPQPMQOYIVLGVTTIVDVIIVMIGVATLA 169
169 ORIALWIKGPKQKALNKFGLSLFMLVGALLASAR 203
170 TRIARWIKSPQMKLNIRFGGLFMLICALLATAR 204

RESULT 9
AAG98975
ID AAG98975 standard; protein; 138 AA.
XX AC AAG98975;
XX DT 26-SEP-2001 (first entry)
XX DE E. coli growth and proliferation related protein sequence SEQ ID NO:445.
XX KW Escherichia coli; growth; proliferation; microbial; antimicrobial;
XX OS bacterial infection; microorganism.
XX PN Escherichia coli.
XX WO200134810-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030950.
XX PR 09-NOV-1999; 99US-0164415P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Forsyth RA, Ohlsen K, Zyskind J;
XX DR WPI; 2001-335933/35.
XX PT N-PSDB; AAH84646.
XX PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
XX for screening for homologous genes and for designing expression vectors.
XX PS Claim 19; Page 502-503; 522pp; English.
XX CC AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
XX related DNA sequences (1). AAH84500 to AAH84670 encode the E. coli growth
XX and proliferation related proteins given in AAG99078 and AAG98830 to
XX AAG98999 (1) can be used as potential targets for the generation of new
XX antimicrobial agents, and for identification of compounds which interact
XX with the gene products of (1). In addition the expression of (1) and the
XX purification of the proteins, the purified proteins can be used to
XX generate reagents and screen small molecule libraries or other candidate
XX compound libraries for compounds that can be further developed to yield
XX novel antimicrobial compounds. In addition, nucleic acid probes
XX complementary to (1) that are specific for particular species of
XX microorganisms can be used to identify particular microorganism species
XX in clinical specimens, therefore, providing a rapid and dependable method
XX by which to identify the causative agents of a bacterial infection. Also,
XX antibodies generated against proteins translated from mRNA transcribed
XX from proliferation-required sequences can also be used to screen for
XX specific microorganisms that produce such proteins in a species-specific
XX manner. AAH84371 and AAH84670 represent sequencing primers used in the
XX isolation of E. coli growth and proliferation related sequence, which are
XX used in an example from the present invention

Sequence 138 AA;
Query Match 64.7%; Score 686; DB 4; Length 138;
Best Local Similarity 99.3%; Pred. No. 4.9e-70;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
68 VTAPEVLKWAAGAYLWLGICQWRAAGADLKSLASTQSRRLHFORAVFNLTNPKSIVF 127
1 MIAPEVLKWAAGAYLWLGICQWRAAGADLKSLASTQSRRLHFORAVFNLTNPKSIVF 60
128 LAALFPQIPQPMQOYIVLGVTTIVDVIIVMIGVATLAORIALWIKGPKQKALNKI 187
61 LAALFPQIPQPMQOYIVLGVTTIVDVIIVMIGVATLAORIALWIKGPKQKALNKI 120
188 FGSFLMVLVGALLASARHA 205
121 FGSFLMVLVGALLASARHA 138

CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: the sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 205 AA;

Query Match 43.3%; Score 459.5; DB 6; Length 205;
Best Local Similarity 43.4%; Pred. No. 6.3e-44;
Matches 89; Conservative 44; Mismatches 69; Indels 3; Gaps 2;
QY 1 MTEWFAVLLTSIIITLSPGSGAINTMTSTLNHG--YPAGGVYCWASDRGDSYCAGWR 58
DB 1 MDTHWMLAYLLTAVVSLAPFGSGTVNSISNGLSYGRSLGALIGLQGLACHIVLVG-I 59
QY 59 GVGTLSRSVIAEVLKWAAGAYLIWLGIOQWRAAGADLKLASTQSRRLHFORAVFVN 118
DB 60 GIGALVAQSAALPTLLKWIAGAYLVWLGIOQWRAAGADLKLASTQSRRLHFORAVFVN 119
QY 119 LTNPKSIVFLAALFPQFIMPOQPMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGP 178
DB 120 LTNPKSIVFLAALFPQFIMPOQPMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGP 179
QY 179 KQWALKNKIFGSLFVLVGLLASAR 203
DB 180 NIMTRNNKIFGSMFGCGMLLATAK 204

RESULT 12
ABG25225
ID ABG25225 standard; protein; 245 AA.
XX AC ABG25225;
XX AC ABG25225;
DT 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #25216.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
KW Homo sapiens.
OS Homo sapiens.
XX W0200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS89412.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 55584; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 245 AA;
Query Match 38.6%; Score 410; DB 4; Length 245;
Best Local Similarity 94.3%; Pred. No. 3.6e-38;
Matches 83; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 110 LQRAVFNLTNPKSIVFLAALFPQFIMPOQPMQYIVLGVTTIVVDIIVMIGYATLAQ 169
DB 8 MYIRAVFNLTNPKSIVFLAALFPQFIMPOQPMQYIVLGVTTIVVDIIVMIGYATLAQ 67
QY 170 RIALWIKGPKQWALKNKIFGSLFVLVGA 197
DB 68 RIALWIKGPKQWALKNKIFGSLFVLVHNA 95
RESULT 13
ABU38890
ID ABU38890 standard; protein; 203 AA.
XX AC ABU38890;
XX AC ABU38890;
DT 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #24417.
DE Protein encoded by Prokaryotic essential gene #24417.
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
KW Pseudomonas aeruginosa.
OS Pseudomonas aeruginosa.
XX W0200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
DR N-PSDB; ACA42760.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 66814; 1766pp; English.
PS
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 209 AA;

Query Match 36.7%; Score 399; DB 6; Length 209;
Best Local Similarity 36.4%; Pred. No. 7.4e-36;
Matches 83; Conservative 37; Mismatches 64; Indels 44; Gaps 4;
QY 1 MTLWWFAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWR-- 58
Db 1 MLVSTWFAFFLACWAISLSPGAGAIASMSGLQYGFARG-----YNNAL 44
QY 59 -----GVGTLSRSVIAFEVLKWAAGAYLIWLGIQWRA-AGAILDKSLASTQSR 103
Db 45 GLQIGLALQIAIVAAGVALLATSAFSLIKWFGVAYLVLAVRQWQAP-----PQALS 99
QY 104 TQSRHLFQ-----RAVFVNLTNPKSVLFAALFPQFIMPOQPMQYIVLGVTTIIVD 157
Db 100 TQGERPLGRPLTVLURGFVNASHPKAVIFMLAVLPQFIDPHQPLLAQYLINGGTMIIVD 159
QY 158 ITVMIGYATLAQRIALWIKPKQKMAINKIFGSLFMLVGALLASARHA 205
Db 160 LIWVAGYTGAAARVLRVLRSPRQQLVNRTPASLFGVGAAGLLATVRA 207

RESULT 14
ADA33616
ID ADA33616 standard; protein; 220 AA.
XX
XX ADA33616;
XX
XX 20-NOV-2003 (first entry)
DT

XX Acinetobacter baumannii protein #777.
DE
XX
XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
KW plant biocontrol agent.
XX
XX Acinetobacter baumannii.
OS
XX
XX US6562958-B1.
PN
XX
XX 13-WAY-2003.
PD
XX
XX 04-JUN-1999; 99US-00328352.
PF
XX
XX 09-JUN-1998; 98US-0088701P.
PR
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX
XX Breton G, Rush D;
PI
XX
XX WPI: 2003-576092/54.
DR
XX
XX N-PSDB; ADA29490.
DR
XX
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
XX Example; SEQ ID NO 4903; 328pp; English.
PS
XX
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents the amino acid sequence of an A.
CC baumannii protein.
XX
XX Sequence 220 AA;
SQ
Query Match 34.7%; Score 368; DB 6; Length 220;
Best Local Similarity 35.3%; Pred. No. 2e-33;
Matches 77; Conservative 71; Indels 24; Gaps 3;
QY 1 MTLWWFAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWR-- 58
Db 13 MSLQWFAFAYMLACWVISISPGAGAIASMSGLNYGF-----RHGYWNAIGLQIA 61
QY 59 -----GVGTLSRSVIAFEVLKWAAGAYLIWLGIQWRA-AGAILDKSLASTQSR 107
Db 62 LLIQIMIVAAGVGVLFATTPLAQAVKWFQVALLVLAQLWTAPVKDIEIQHEKKDSV 121
QY 108 RHLFQRAVFVNLTNPKSVLFAALFPQFIMPOQPMQYIVLGVTTIIVDITVMIGYATL 167
Db 122 SALLFNFGVWNISNPKAIVELLAVLPQFLDLSKFPQWIOYLIQAATMTIDLIWMAGYTGL 181
QY 168 AQRLATWIKPKQKMAINKIFGSLFMLVGALLASARHA 205
Db 182 ASKVLRLIRSPKQKQYLNRGFAYNFCALLLSVTHQA 219
RESULT 15
ABU41414
ID ABU41414 standard; protein; 210 AA.
XX
XX AC ABU41414;
XX
XX 19-JUN-2003 (first entry)
XX
XX DE Protein encoded by prokaryotic essential gene #26941.
XX

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Pseudomonas syringae.
OS WO200277183-A2.
PN PD 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342323P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA45284.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 69338; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 210 AA;
SQ
Query Match 34.6%; Score 367.5; DB 6; Length 210;
Best Local Similarity 34.5%; Pred. No. 2.1e-33;
Matches 76; Conservative 40; Mismatches 69; Indels 35; Gaps 3;
QY 1 MTELEWFAVLLTSLITLSPGSGAINTMTTSINHGYPAGGVYCWASDRGTGDSYCAQWR-- 58
DB 1 MALETWLGFFAACWIIISLSPGAGAIASMSGGLQYGLFRG-----YNNAL 44
QY 59 -----GVGTFLFSRVIATFEVLKWAGAYLIMLGIQWRAA--GAIDLKSL 101

Db 45 GLQIALVAQIAVAAGLGAALAAASEMAFTLIKWFVAYLVYLGIKQWRASPTDLADESAV 104
QY 102 ASTQSRRLHFORAVFVNLTNPKSIVFLAALFPQFIMPQQPOLMQYIVLGVYTTIWDIIIVM 161
Db 105 RPVGKPMTLVPRGFLVNISNPKALIFILALPQFIETPAELFMQYVYIIAATWVVVDLIIVM 164
QY 162 IGYATLAQRIALTIKQPKQMKALNKIPGSLFMLVGGALLAS 201
Db 165 AGYTGGLASKVLRLAKTTPQQRRLNRTFASLFVGAAGFLAT 204

Search completed: March 24, 2004, 21:14:43
Job time : 53 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:13:32 ; Search time 22 Seconds
(without alignments)
481.060 Million cell updates/sec

Title: US-09-847-392-2

Perfect score: 1061

Sequence: 1 MLEWFAVLLTSIIITLSP.....KIFGSLFVLGALLASARHA 205

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCUS-COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1061	100.0	205	4	US-09-396-357-2
2	832.5	78.5	214	4	US-09-489-039A-8066
3	686	64.7	138	4	US-09-711-164-445
4	389	36.7	277	4	US-09-352-991A-23018
5	368	34.7	220	4	US-09-328-352-4903
6	228	21.5	250	4	US-09-543-681A-4487
7	223	21.0	210	4	US-09-328-352-6808
8	215.5	20.3	241	4	US-09-328-352-7159
9	200.5	18.9	214	4	US-09-489-039A-9793
10	198	18.7	208	4	US-09-328-352-5111
11	193	18.2	219	4	US-09-489-039A-7731
12	193	18.2	222	4	US-09-352-991A-26644
13	179.5	16.9	237	4	US-09-352-991A-27175
14	175.5	16.5	211	4	US-09-352-991A-20747
15	170	16.0	240	4	US-09-328-352-7537
16	168.5	15.9	235	4	US-09-328-352-8249
17	164.5	15.5	228	4	US-09-543-681A-4854
18	163.5	15.4	212	4	US-09-543-681A-4767
19	156	14.7	216	4	US-09-352-991A-18666
20	155	14.6	153	2	US-08-476-254-7
21	155	14.6	153	6	5474933-4
22	148	13.9	226	4	US-09-352-991A-31610
23	147.5	13.9	211	4	US-09-489-039A-9608
24	147	13.9	249	4	US-09-328-352-5813
25	144	13.6	217	4	US-09-489-039A-8076
26	140	13.2	260	4	US-09-352-991A-25992
27	138.5	13.1	220	4	US-09-543-681A-4545

ALIGNMENTS

RESULT 1

US-09-396-357-2

; Sequence 2, Application US/09396357

; Patent No. 6303348

; GENERAL INFORMATION:

; APPLICANT: LIVSHITS, VITALY ARKADIEVICH

; APPLICANT: ZAKHAEVA, NATALIYA PAVLOVNA

; APPLICANT: ALCOSHIN, VLADIMIR VENIAMOVICH

; APPLICANT: BELAREOVA, ALL VALENTINOVNA

; APPLICANT: TOKHAKOVA, IRINA LIOVNA

; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONPERS ON BACTERIUM

; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD

; FILE REFERENCE: 0010-1039-0

; CURRENT APPLICATION NUMBER: US/09/396,357

; CURRENT FILING DATE: 1999-09-15

; EARLIER APPLICATION NUMBER: RU98118425

; EARLIER FILING DATE: 1998-10-13

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 205

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-396-357-2

Query Match 100.0%; Score 1061; DB 4; Length 205;

Best Local Similarity 100.0%; Pred. No. 1,1e-116; Indels 0; Gaps 0;

Matches 205; Conservative 0; Mismatches 0;

Qy 1 MLEWFAVLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYVCWASDRDTSYCAWGV 60

Db 1 MLEWFAVLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYVCWASDRDTSYCAWGV 60

Qy 61 GTLFSSVTAFLVKWAGAAVILWGIQWRAAGDAIDKSLASTQSRHLPQRAVFNLT 120

Db 61 GTLFSSVTAFLVKWAGAAVILWGIQWRAAGDAIDKSLASTQSRHLPQRAVFNLT 120

Qy 121 NPKSIVFLAALFPQFMQYIVLVGTVTVIWDIIVMIGVATLAQRIALWIKGPKQ 180

Db 121 NPKSIVFLAALFPQFMQYIVLVGTVTVIWDIIVMIGVATLAQRIALWIKGPKQ 180

Qy 181 MKALNKIFGSLFVLGALLASARHA 205

Db 181 MKALNKIFGSLFVLGALLASARHA 205

RESULT 2

US-09-489-039A-8066

; Sequence 8066, Application US/09489039A

Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 8066
LENGTH: 214
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8066

Query Match 78.5%; Score 832.5; DB 4; Length 214;
Best Local Similarity 81.7%; Pred. No. 8.9e-90;
Matches 170; Conservative 12; Mismatches 21; Indels 5; Gaps 3;

QY 1 MTEWFWAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRGTG--DSCAGW 57
DB 9 MTEWFWAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRGTG--DSCAGW 56

QY 58 RGVGLFGRSVIAFEVLKWAAGAAVLIMIGIQWRAAGAILDKSLASTQSRHLPQRAVVF 117
DB 67 VGLGLTGRSVIAFEVLKWAAGAAVLIMIGIQWRAAGAILDKSLASTQSRHLPQRAVVF 126

QY 118 NLTNPKSVIFLAALFPQIMQPQOLQYIVLVGTTTVDIIVMIGYATLAQRIALWIKG 177
DB 127 NLTNPKSVIFLAALFPQIMQPQOLQYIVLVGTTTVDIIVMIGYATLAQRIALWIKG 186

QY 178 PQMKALNKIFGSLFMLVGALLASARHA 205
DB 187 PQMKALNKIFGSLFMLVGALLASARHA 214

RESULT 3
US-09-711-164-445
Sequence 445, Application US/09711164
Patent No. 6589738
GENERAL INFORMATION:
APPLICANT: Forsyth, R. Allyn
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
FILE REFERENCE: ELITRA.008A
CURRENT APPLICATION NUMBER: US/09/711,164
PRIOR FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: US 60/164415
PRIOR FILING DATE: 1999-11-9
NUMBER OF SEQ ID NOS: 469
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 445
LENGTH: 138
TYPE: PRT
ORGANISM: Escherichia coli
US-09-711-164-445

Query Match 64.7%; Score 686; DB 4; Length 138;
Best Local Similarity 99.3%; Pred. No. 8.1e-73;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VTAFEVLKWAAGAAVLIMIGIQWRAAGAILDKSLASTQSRHLPQRAVFNLTNPKSVIF 127
DB 1 MTAFEVLKWAAGAAVLIMIGIQWRAAGAILDKSLASTQSRHLPQRAVFNLTNPKSVIF 60

QY 128 LAALFPQIMQPQOLQYIVLVGTTTVDIIVMIGYATLAQRIALWIKGPKOMKALNKI 187
DB 61 LAALFPQIMQPQOLQYIVLVGTTTVDIIVMIGYATLAQRIALWIKGPKOMKALNKI 120

QY 188 FGSLFMLVGALLASARHA 205

Db 121 FGSLFMLVGALLASARHA 138

RESULT 4
US-09-252-991A-23018
Sequence 23018, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23018
LENGTH: 277
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23018

Query Match 36.7%; Score 389; DB 4; Length 277;
Best Local Similarity 36.4%; Pred. No. 1.8e-37;
Matches 83; Conservative 37; Mismatches 64; Indels 44; Gaps 4;

QY 1 MTEWFWAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRGTGSDYCAGWR-- 58
DB 69 MLVSTWTFAPFLACWALSISPGAGAIASMSGQYGFARG-----YWNAL 112

QY 59 -----GVGTLFGRSVIAFEVLKWAAGAAVLIMIGIQWRAAGAILDKSLAS 103
DB 113 GLQIGLALQIAIVAAGVGALLATSALAFSLIKWFGVAYLVYLAVRQWQAP-----PQALS 167

QY 104 TQSRHLPQ-----RAVFNLTNPKSVIFLAALFPQIMQPQOLQYIVLVGTTTVD 157
DB 168 TDGERPLGRPLTLVLRGFLVNASNPKAVIFMLAVLPQFIDPHQLLAQYLIMGTMIYVD 227

QY 158 IIVMIGYATLAQRIALWIKGPKOMKALNKIFGSLFMLVGALLASARHA 205
DB 228 LIIVAGYTGAAARVLRVRSRPROCKLVNRTFASLFVGAAGLLATVRA 275

RESULT 5
US-09-328-352-4903
Sequence 4903, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4903
LENGTH: 220
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4903

Query Match 34.7%; Score 368; DB 4; Length 220;
Best Local Similarity 35.3%; Pred. No. 3.8e-35;
Matches 77; Conservative 46; Mismatches 71; Indels 24; Gaps 3;

QY 1 MTEWFWAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRGTGSDYCAGWR-- 58
DB 13 MSLQWFWAYMLACWVISISPGAGAIASMSGSLNYGF-----RHGYWNAIGLQIA 61

QY 59 -----GVCGLFSSRSVIAFEVLKWAAGAYLWLGICQWFA-AGAIIDLKSLASTQSR 107
DB 62 LLIQIMIVAAGVGLFATTTPLAFQAVKGFVAYLLAYLQWTPVKDIEIQHEKKDSV 121
QY 108 RHLFQRAVFNLTNPKSIVFLAALFPQFIMPOQPMQVILGVTTIVVDIIVMIGYATL 167
DB 122 SALLFNGFVNINPKAIVFLLAVLPQFLDLSXPQIMQVILMAATWTDLIVMAGYTGL 181
QY 168 AORTALWIKGPKOMKALNKFGLSLFMLVGLALLASARHA 205
DB 192 ASKVLRLRLRSKQOYLNRGFAVNFSCAALLLSTVHQHA 219

RESULT 6
US-09-543-681A-4487
; Sequence 4487, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 487
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4487

Query Match 21.5%; Score 228; DB 4; Length 250;
Best Local Similarity 30.8%; Pred. No. 1,3e-18;
Matches 69; Conservative 41; Mismatches 62; Indels 52; Gaps 9;

QY 7 FAYLLTSIILTLSPGSAINTMTSLNHYGYPAG-----GVYCWASDRTGDSYC--AGWRG 59
DB 46 WTYLLGVIFITLVPGPSIFVLTSKAGHKGKGYKALGVF-----TGDALLIFLAFLG 99
QY 60 VGTLFSSRSVIAFEVLKWAAGAYLWLGICQWRAAGALDKSLAST----- 104
DB 100 VASLVKTSVPFFVIKAGALYLYLG-----LKLTYATHPKKKEQEQVAVEV 148
QY 105 -QSRRLHFORAVFNLTNPKSIVFLAALFPQFIMPOQPMQVILGV-----TTIIVVD 157
DB 149 VKXNGLYIKALFLSLNPKMIFVVSFFTFIDPKVENAGVFPFVLGVILEICSMYLS 208
QY 158 IIVMIGYATLAQRTALWIKGPKOMKAL-NKIFGSLFMLVGLALLA 200
DB 209 VLIFFGVA-----ITNKVGNKRLASLSNSCIGAVFLLFGAKLA 247

RESULT 7
US-09-328-352-6808
; Sequence 6808, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: GARY L. BRETON et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6808
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6808

Query Match 21.0%; Score 223; DB 4; Length 210;

Best Local Similarity 30.0%; Pred. No. 4e-18;
Matches 61; Conservative 38; Mismatches 82; Indels 22; Gaps 5;

QY 11 LTSIILTL-SPGSAINTMTSLNHYGYPAG-----GVYCWASDRTGDSYCAGWRGV 60
DB 13 ITIAFLTLSPGPGVLPFTVNTGYNVGTALFGISGLIIGMEFIA-----VISASGV 64
QY 61 GTLFSSRSVIAFEVLKWAAGAYLWLGICQW-PAAGAIIDLKSLASTQ-----SRRLHFORAVE 116
DB 65 GLIITSNPTIFALTKEIGAFYLMYLGYNFKIKKSTQDLETEHNEKIGKSKLFYQGLF 124
QY 117 VNLTPKSIIVFAALFPQFIMPOQPMQVILGVTTIVVDIIVMIGYATLAQRTALWIK 176
DB 125 ASLNPKTIVFPFIALPQFIDIKKEILNQFLVLSLTFCLIGFLIHLVYANPSSIFKEMKL 184
QY 177 GPKOMKALNKFGLSLFMLVGLALL 199
DB 185 AGNFKLNKVGSCIFFLAVILL 207

RESULT 8
US-09-328-352-7159
; Sequence 7159, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: GARY L. BRETON et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7159
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7159

Query Match 20.3%; Score 215.5; DB 4; Length 241;
Best Local Similarity 28.5%; Pred. No. 3.7e-17;
Matches 65; Conservative 45; Mismatches 77; Indels 41; Gaps 10;

QY 2 TLEWFA-----YLLTSIILTLSPGSAINTMTSLNHYG-----YPAGGVYCWASDR 49
DB 28 SIIRWTYMTAILPYLIAITLTLTLPGLDTHIIRTALEGSKAFQALGI----- 79
QY 50 GDSYCAGWR-----GVGTLFSSRSVIAFEVLKWAAGAYLWLGICQ-----QWRAAGAIIDLK 99
DB 80 -SLGCIANGIIVWACGLGALLMASDLAFNLKWMGAAYLAWGLNMLKPRSQLADIQDNH 138
QY 100 SLASTQSRRLHFORAVFNLTNPKSIVFLAALFPQFIMPOQPMQVILGVTTIVVDII 159
DB 139 SNRSTS--ENWFIKGFGLNLPKVGIFVIFLFPQFI-PAQASAVTW-VWGL--VNIHV 192
QY 160 VMIQYATL-----AORTALWIKGPKOMKALNKFGLSLFMLVGLALLASAR 203
DB 193 IGVLSLSLLILANQPLSRYLKQPKFYKYMDRITGSIFVLFAKLAFSK 240

RESULT 9
US-09-489-039A-9793
; Sequence 9793, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:16:17 ; Search time 41 Seconds

(without alignments)

1294.777 Million cell updates/sec

Title: US-09-847-392-2

Perfect score: 1061

Sequence: 1 MTLWEFAYLLTSILTLSP.....KIRGSLFVLGALLASARHA 205

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpa/FCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpa/FCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1061	100.0	205	9	US-09-927-395-2
2	1061	100.0	205	9	US-09-847-392-2
3	830.5	78.3	206	12	US-10-282-122A-60157
4	827.5	78.0	206	12	US-10-282-122A-56211
5	818.5	77.1	206	12	US-10-282-122A-75782
6	791	74.6	205	12	US-10-282-122A-72863
7	698.5	65.8	206	12	US-10-282-122A-78024
8	586	64.7	138	12	US-10-282-122A-42680
9	586	64.7	138	14	US-10-287-274-445
10	459.5	43.3	205	12	US-10-282-122A-76964
11	389	36.7	209	12	US-10-282-122A-66814
12	367.5	34.6	210	12	US-10-282-122A-69338
13	354.5	33.4	211	12	US-10-282-122A-67883
14	312.5	29.5	209	12	US-10-282-122A-51102
15	291	27.4	181	12	US-10-282-122A-44873

16	290.5	27.4	211	12	US-10-282-122A-47826	Sequence 47826, A
17	180.5	17.0	217	14	US-10-156-761-18918	Sequence 10318, A
18	163.5	15.4	212	12	US-10-374-903A-6	Sequence 6, Appli
19	150	14.1	207	9	US-09-738-626-6418	Sequence 6418, Ap
20	147	13.9	224	14	US-10-156-761-7796	Sequence 7796, Ap
21	139.5	13.1	205	14	US-10-156-761-11297	Sequence 11297, A
22	134.5	12.7	226	9	US-09-738-626-3665	Sequence 3665, Ap
23	126	11.9	195	12	US-10-620-487-2	Sequence 2, Appli
24	105.5	9.9	223	9	US-09-738-626-6070	Sequence 6070, Ap
25	105.5	9.9	223	10	US-09-746-660A-14	Sequence 14, Appl
26	94	8.9	456	12	US-10-282-122A-75400	Sequence 75400, A
27	94	8.9	457	12	US-10-282-122A-43084	Sequence 43084, A
28	94	8.9	457	15	US-10-369-493-729	Sequence 729, App
29	94	8.9	463	9	US-09-815-242-13918	Sequence 13918, A
30	93	8.8	448	12	US-10-282-122A-72853	Sequence 72853, A
31	92.5	8.7	363	12	US-10-282-122A-46325	Sequence 46325, A
32	92	8.7	456	12	US-10-282-122A-74836	Sequence 74836, A
33	91	8.6	457	12	US-10-282-122A-60097	Sequence 60097, A
34	91	8.6	573	12	US-10-425-114-67027	Sequence 67027, A
35	89.5	8.4	233	10	US-09-746-660A-52	Sequence 52, Appl
36	89.5	8.4	233	14	US-10-196-232-25	Sequence 25, Appl
37	89.5	8.4	236	9	US-09-738-626-6955	Sequence 6955, Ap
38	89.5	8.4	236	14	US-10-166-142-8	Sequence 8, Appli
39	87.5	8.2	465	12	US-10-282-122A-43620	Sequence 43620, A
40	87	8.2	364	12	US-10-424-599-247108	Sequence 247108,
41	85	8.0	484	9	US-09-815-242-10809	Sequence 10809, A
42	85	8.0	484	12	US-10-282-122A-56917	Sequence 56917, A
43	83.5	7.9	438	12	US-10-282-122A-61182	Sequence 61182, A
44	83.5	7.9	506	12	US-10-282-122A-49913	Sequence 49913, A
45	83.5	7.9	509	12	US-10-282-122A-49320	Sequence 49320, A

ALIGNMENTS

RESULT 1

US-09-927-395-2
; Sequence 2, Application US/09927395
; Patent No. US20020058314A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENYAMOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHAKOVA, IRINA LVOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM RESISTANCE TO L0-HOMOSERINE AND METHOD FOR PRODUCING L-AMINO ACIDS
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/927,395
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-927-395-2

Query Match 100.0%; Score 1061; DB 9; Length 205;
Best Local Similarity 100.0%; Pred No. 6.2e-110;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTLWEFAYLLTSILTLSPGCAINTMTLSLNGYPAGGVTCWASDRDTSYCAWRGV 60

Db 1 MTLWEFAYLLTSILTLSPGCAINTMTLSLNGYPAGGVTCWASDRDTSYCAWRGV 60

QY 61 GTLPSRVIAFEVLKWAAYLLWLGLCQWRAGATDLKSLASTQSRRLFORAVFVNL 120

Db 61 GTLFSRSVAFEVKWAAGAYLWLGIOQWRAAGAILDKSLASTQSRRLHFORAVFNLT 120
QY 121 NPKSIVFLAALFPQFIMPQPOLMQLVGLVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180
Db 121 NPKSIVFLAALFPQFIMPQPOLMQLVGLVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180
QY 181 MKALNKIFGSLFVLGALLASARHA 205
Db 181 MKALNKIFGSLFVLGALLASARHA 205

RESULT 2

US-09-847-392-2
; Sequence 2, Application US/09847392
; Patent No. US20020102670A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENAMIOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHAKOVA, IRINA LIOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/847,392
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-847-392-2

Query Match 100.0%; Score 1061; DB 9; Length 205;
Best Local Similarity 100.0%; Pred. No. 6.2e-110; Indels 0; Gaps 0;
Matches 205; Conservative 0; Mismatches 0;
QY 1 MLEWFWAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGYCWSRDTGDSYCAWRGV 60
Db 1 MLEWFWAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGYCWSRDTGDSYCAWRGV 60
QY 61 GTLFSRSVAFEVKWAAGAYLWLGIOQWRAAGAILDKSLASTQSRRLHFORAVFNLT 120
Db 61 GTLFSRSVAFEVKWAAGAYLWLGIOQWRAAGAILDKSLASTQSRRLHFORAVFNLT 120
QY 121 NPKSIVFLAALFPQFIMPQPOLMQLVGLVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180
Db 121 NPKSIVFLAALFPQFIMPQPOLMQLVGLVTTIVVDIIVMIGYATLAQRIALWIKGPKQ 180
QY 181 MKALNKIFGSLFVLGALLASARHA 205
Db 181 MKALNKIFGSLFVLGALLASARHA 205

RESULT 3

US-10-282-122A-60157
; Sequence 60157, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60157
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60157

Query Match 78.3%; Score 830.5; DB 12; Length 206;
Best Local Similarity 81.2%; Pred. No. 3.2e-84;
Matches 169; Conservative 13; Mismatches 21; Indels 5; Gaps 3;
QY 1 MLEWFWAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGYCWSRDTG---DSYCAW 57
Db 1 MLEWFWAYLLTSIILTLSPGSGAINTMTTSLNHGYPAGGYCWSRDTG---DSYCAW 57
QY 58 RGVTLSRSVIAFEVLKWAAGAYLWLGIOQWRAAGAILDKSLASTQSRRLHFORAVFV 117
Db 59 VGLGTLFSRSVIAFEVLKWAAGAYLWLGIOQWRAAGAILDKSLASTQSRRLHFORAVFV 118
QY 118 NLTPKSIIVFLAALFPQFIMPQPOLMQLVGLVTTIVVDIIVMIGYATLAQRIALWIKG 177
Db 119 NLTPKSIIVFLAALFPQFIMPQPOLMQLVGLVTTIVVDIIVMIGYATLAQRIALWIKG 178
QY 178 PKMKALNKIFGSLFVLGALLASARHA 205
Db 179 PKMKALNKIFGSLFVLGALLASARHA 206

RESULT 4

US-10-282-122A-56211
; Sequence 56211, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-30
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 56211
LENGTH: 206
TYPE: PRT
ORGANISM: Enterobacter cloacae
US-10-282-122A-56211

Query Match 78.0%; Score 827.5; DB 12; Length 206;
Best Local Similarity 79.1%; Pred. No. 7e-84;
Matches 167; Conservative 15; Mismatches 18; Indels 11; Gaps 3;
Qy 1 MTLBWFAYLLTSLTSPGSGAINTMTTSLNHGYPAGGVYCWASDRGTGDSYACAGWR-- 58
Db 1 MTFEWFAYLLTSLTSLSPGSGAINTMTTSLNHGY-RGAAASIAGLQTG---LGIHIV 55
Qy 59 ---GVGTFLFSRSVIAFEVLKWAAGAYLIWLGTCQWRAAGAILDKSLASTQSRHLFORA 114
Db 56 LVGIGLGTFLFSRSVIAFEVLKWAAGAYLIWLGTCQWRAAGSINLNTALTQNRHLFKRA 115
Qy 115 VFVNLTPKSIIVFLAALFPQFIMPOQPOLMOYVLGVTTIIVDDIIVMIGYATLAQRIALW 174
Db 116 VFVNLTPKSIIVFLAALFPQFIVPHQPNQYVVLGATTIIVDDIIVMIGYATLAQRIAAW 175
Qy 175 IKGPKQKALKNKIFGSLFMLVGALLASARHA 205
Db 176 IKGPKQKALKNKVFGSLFMLVGALLASARHA 206

RESULT 5
US-10-282-122A-75782
Sequence 75782, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-30
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 75782
LENGTH: 206
TYPE: PRT
ORGANISM: Salmonella typhi
US-10-282-122A-75782

Query Match 77.1%; Score 819.5; DB 12; Length 206;
Best Local Similarity 78.7%; Pred. No. 7e-83;
Matches 166; Conservative 13; Mismatches 21; Indels 11; Gaps 3;
Qy 1 MTLBWFAYLLTSLTSPGSGAINTMTTSLNHGYPAGGVYCWASDRGTGDSYACAGWR-- 58
Db 1 MTFEWFAYLLTSLTSLSPGSGAINTMTTSLNHGY-RGAAASIAGLQTG---LGIHIV 55
Qy 59 ---GVGTFLFSRSVIAFEVLKWAAGAYLIWLGTCQWRAAGAILDKSLASTQSRHLFORA 114
Db 56 LVGIGLGTFLFSRSVIAFEVLKWAAGAYLIWLGTCQWRAAGAILDLHTLAQTSGRGLFKRA 115
Qy 115 VFVNLTPKSIIVFLAALFPQFIMPOQPOLMOYVLGVTTIIVDDIIVMIGYATLAQRIALW 174
Db 116 VFVNLTPKSIIVFLAALFPQFIMPOQPOLAQLIILGVTTIIVDDIIVMIGYATLAQRIAAW 175
Qy 175 IKGPKQKALKNKIFGSLFMLVGALLASARHA 205
Db 176 IKGPKQKALKNKAFGSLFMLVGALLASARHA 206

RESULT 6
US-10-282-122A-72863
Sequence 72863, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72863
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (169)..(169)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-72863

Query Match 74.6%; Score 791; DB 12; Length 205;
Best Local Similarity 76.8%; Pred. No. 8.1e-80;
Matches 162; Conservative 14; Mismatches 23; Indels 12; Gaps 4;
QY 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWR-- 58
DB 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWR-- 58
QY 59 ----GVGTLFERSVIAFEVLKWAGAAVLIWLGIOQWRAGAAIDLKSLASTQSRHIFQRA 114
DB 56 LVGVGLGTLFERSVIAFEVLKWAGAAVLIWLGIOQWRAGAAIDLKSLASTQSRHIFQRA 114
QY 115 VFVNLNPKSIVFLAALPQFIMPQQLMXYIVLGVTTIVVDIIVMIGVATLAQRITAW 174
DB 115 IFVNLNPKSIVFLAALPQFIMPQQLMXYIVLGVTTIVVDIIVMIGVATLAQRITAW 174
QY 175 IKGPQKMKALNPKFQSLFMLVGALLASARHA 205
DB 175 IKGPQKMKALNPKFQSLFMLVGALLASARHA 205

RESULT 7
US-10-282-122A-78024
; Sequence 78024, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2003-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78024
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-78024

Query Match 65.8%; Score 698.5; DB 12; Length 206;
Best Local Similarity 63.3%; Pred. No. 1.7e-89;
Matches 136; Conservative 33; Mismatches 23; Indels 23; Gaps 3;
QY 1 MTEWFWAYLLTSIIITLSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWR-- 58
DB 1 MTDLWLTLLTLLTLLSLSFGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGWR-- 58
QY 59 -----GVGTLFERSVIAFEVLKWAGAAVLIWLGIOQWRAGAAIDLKSLASTQSR 108
DB 50 LAVHIVLVGVLGALVQSLLAFEILKWLGAAYLIWLGIOQWRAGAAIDLKSLASTQSR 109
QY 109 HLFQRAVFNLTNPKSIVFLAALPQFIMPQQLMXYIVLGVTTIVVDIIVMIGVATLA 168
DB 110 KLFQRAVFNLTNPKSIVFLAALPQFIMPQQLMXYIVLGVTTIVVDIIVMIGVATLA 169
QY 169 QRIALWIKGPKMKALNPKFQSLFMLVGALLASAR 203
DB 170 TRIARWIKSPQMKLNRIFGGLFMLIGALLATAR 204

RESULT 8
US-10-282-122A-42680
; Sequence 42680, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A

```

; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42660
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-42660

Query Match
Best Local Similarity 54.7%; Score 686; DB 12; Length 138;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQSRHLFQRAVFNLTNPKSIYF 127
Db 1 MIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQSRHLFQRAVFNLTNPKSIYF 60

QY 128 LAALFPQIMFQPOLMOYIVLGVTTIIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 187
Db 61 LAALFPQIMFQPOLMOYIVLGVTTIIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 120

QY 188 FGSLEMLVGALLASARHA 205
Db 121 FGSLEMLVGALLASARHA 138

RESULT 9
US-10-287-274-445
; Sequence 445, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETO
; FILE REFERENCE: ELITRA.008DV1
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 445
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-287-274-445

Query Match
Best Local Similarity 64.7%; Score 686; DB 14; Length 138;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQSRHLFQRAVFNLTNPKSIYF 127
Db 1 MIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQSRHLFQRAVFNLTNPKSIYF 60

QY 128 LAALFPQIMFQPOLMOYIVLGVTTIIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 187
Db 61 LAALFPQIMFQPOLMOYIVLGVTTIIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 120

QY 188 FGSLEMLVGALLASARHA 205
Db 121 FGSLEMLVGALLASARHA 138
```

```

Best Local Similarity 99.3%; Pred. No. 2.5e-68;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQSRHLFQRAVFNLTNPKSIYF 127
Db 1 MIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQSRHLFQRAVFNLTNPKSIYF 60

QY 128 LAALFPQIMFQPOLMOYIVLGVTTIIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 187
Db 61 LAALFPQIMFQPOLMOYIVLGVTTIIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 120

QY 188 FGSLEMLVGALLASARHA 205
Db 121 FGSLEMLVGALLASARHA 138

RESULT 10
US-10-282-122A-76964
; Sequence 76964, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76964
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-76964

Query Match
Best Local Similarity 43.3%; Score 459.5; DB 12; Length 205;
Matches 89; Conservative 44; Mismatches 69; Indels 3; Gaps 2;

QY 1 MTLFWFAYLLTIIILTSFGSGAINTMTSLNHG--YPAGGYVCWASRTGDSYCAGR 58
Db 1 MDIHWLAVLLTAVFSLAPGSGTVNSINGLSYGTRHSLGILGIGLACHIVLVG-I 59
```

QY 59 GVTTLFSRSVIAFEVLKWAGAYLIWLGIQOWRAAGAILDKSLASTOSRRHLFORAVFVN 118
Db 60 GTGALVAQSALAEFTLKTKTGAAYLVMLGIQKWRDRAPLTATTTSHELSQAALLRKAVLIN 119
QY 119 LTNPKSIVFLAALFPQFIMPOQPMQVIVLGVTTIVVDIIVMIGYATLAORALMWIKGP 178
Db 120 LTNPKSIVFLAALFPQFIDPTDHPQFVLVGIITVTDAIVMGYALAAQLGRYIRSP 179
QY 179 KQKALNKIFGSLFMLVGLASAR 203
Db 180 NIMTMNKLFSGSMFGCGMLLATAK 204

RESULT 11
US-10-282-122A-66814
; Sequence 66814, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 66814
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66814

Query Match 36.7%; Score 389; DB 12; Length 209;
Best Local Similarity 36.4%; Pred. No. 5.7e-35;
Matches 83; Conservative 37; Mismatches 64; Indels 44; Gaps 4;
QY 1 MTEWFWAYLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGWR-- 58
Db 1 MLVSTWFAFELACWATSLSPGAGATASMSGCLQYGFARG-----YNNAL 44
QY 59 -----GVGTLFSRSVIAFEVLKWAGAYLIWLGIQOWRAAGAILDKSLAS 103

Db 45 GLQIGLALQIAIVAGVALLATSAFSLIKKFGVAYVYLAVRQWAP-----PQALS 99
QY 104 TQSRHILFQ-----RAVFNLTNPKSIVFLAALFPQFIMPOQPMQVIVLGVTTIVVD 157
Db 100 TDGSRPLGRPLTLVLRGFLVNASPKAVIFMLAVLPQFIDHPQLLAQYLINGGTIVVD 159
QY 158 IIVMIGYATLAORALMWIKGPQKMKALNKIFGSLFMLVGLALLASARHA 205
Db 160 LIVMAGYTGLAARVLRVRSRQOKLVNRTFASLFGAAGLLATVRA 207

RESULT 12
US-10-282-122A-69338
; Sequence 69338, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 69338
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69338

Query Match 34.6%; Score 367.5; DB 12; Length 210;
Best Local Similarity 34.5%; Pred. No. 1.4e-32;
Matches 76; Conservative 40; Mismatches 69; Indels 35; Gaps 3;
QY 1 MTEWFWAYLTSIIITLSPGSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGWR-- 58
Db 1 MALETWLGFFAACWIIISLSPGAGATASMSGCLQYGFARG-----YNNAL 44
QY 59 -----GVGTLFSRSVIAFEVLKWAGAYLIWLGIQOWRAA--GAIDLKSL 101
Db 45 GLQIALVAQIAIVAGVALLATSAFSLIKKFGVAYVYLAVRQWASPTDLADESAV 104

QY 102 ASTOSRRHLFORAVFVNLTNPKSIVFLAALFPQFIMPOQPOLMQYIVLGVTTIWDIIVM 161
DB 105 RPVGKPTLVFRGLVFNISPKALIFILALPQFIEPTAPLEMQYVIAATVWVDLIWM 164
QY 162 IGYATLAQRIALWIKGPKOMKANKIFGSLFMLVGALLAS 201
DB 165 AGYGLASKVLRAKTPROQRRLNRTFASLFVGNAGFLAT 204

RESULT 13

US-10-282-122A-67883
; Sequence 67883, Application US/10282122A
; Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA 034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: Patent in version 3.1

SEQ ID NO 67883

LENGTH: 211

TYPE: PRT

ORGANISM: Pseudomonas putida

US-10-282-122A-67883

Query Match 33.4%; Score 354.5; DB 12; Length 211;
Best Local Similarity 32.3%; Pred. No. 4.1e-31;
Matches 74; Conservative 41; Mismatches 69; Indels 45; Gaps 4;

QY 1 MTLFWFAYLLTSIIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGWR-- 58

DB 2 MSMEVLGFFAACNVISLSPGAGAIASMSGLOYGF-----WRGY 41

QY 59 -----GVGTLSRSVIAFEVLKWAGAAYLIMLGIOQWRAAGAILDK 99

DB 42 WNALGLQLGLIMQIAITAAAGVAVLASATAFQVWKFGVGLVILAYKQWAL-PMDS 100

QY 100 SLASTQ-----SRRLHFORAVFVNLTNPKSIVFLAALFPQFIMPOQPOLMQYIVLGVTTIIV 156

DB 101 DESGVPIGKPLSLVFRGLVFNISPKALVFLVLPQFINPHALLPQYVAITVTMTV 160
QY 157 DIITWIGYATLAQRIALWIKGPKOMKANKIFGSLFMLVGALLASARHA 205
DB 161 DLLVMAGYGLASHVLRLMLRTPKQKRLNRTFAGLFIGAATFLATLIRRA 209

RESULT 14

US-10-282-122A-51102

; Sequence 51102, Application US/10282122A

; Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA 034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: Patent in version 3.1

SEQ ID NO 51102

LENGTH: 209

TYPE: PRT

ORGANISM: Bordetella pertussis

US-10-282-122A-51102

Query Match 29.5%; Score 312.5; DB 12; Length 209;
Best Local Similarity 33.5%; Pred. No. 2e-26;
Matches 73; Conservative 39; Mismatches 79; Indels 27; Gaps 5;

QY 1 MTLFWFAYLLTSIIILTLSPGSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGW--- 57

DB 1 MTLSTWLTFTFVASWATSFPGAGAIASMSGGUKYGFARG---YWNT-----AGLIIG 49

QY 58 -----RGVGTLSRSVIAFEVLKWAGAAYLIMLGIOQWRAAGA---IDLKSLASTQ 105

DB 50 ILFQFLIVAVGLGAVLATSELAFTLVKYGAVVLYLVGMQRQIRTDAAAPVTVDAGD-PHRA 108

QY 106 SRRLHFORAVFVNLTNPKSIVFLAALFPQFIMPOQPOLMQYIVLGVTTIIVMIGYA 165

DB 109 SIRELVGRGLINTMNPCKGTVFLAVVFPQVDFPQFLTQQYLALAGTLAFTDLVANDVYT 168

QY 166 TLAQRTALMIKPKOMKALNKIFGSLFEMLVGALLASAR 203
Db 169 LLAQVLRMLRKAHHRWNRVFGSLFIAGVFLATER 206

Db 170 AALLISTVHQA 180

Search completed: March 24, 2004, 21:21:52
Job time : 42 secs

RESULT 15

US-10-282-122A-44873
; Sequence 44873, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44873
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-10-282-122A-44873

Query Match 27.4%; Score 291; DB 12; Length 181;
Best Local Similarity 33.5%; Pred. No. 4.1e-24;
Matches 64; Conservative 37; Mismatches 66; Indels 24; Gaps 3;

QY 28 MTTSLNHGYPAGGVYCWASDRGDSYCAGWR-----GVGTLFGRSVIAFEVLK 75
Db 1 MSSGLNYGF-----RHGYWNAIGLQALLIQIMIVAAGVGVLFTTFLAFCAVK 49
QY 76 WAGAAYLLWLGIQWRA-AGADLKLASTOSRRHLFQRAVFNLTNPXSTIVFLAALFPQ 134
Db 50 WFGVAYLLYLVLTQWTPVKDIEIOHEKKDSVALLNGFVNVNINPKRAIVFLAVLPQ 109
QY 135 FIMPOQQLMQVIGLVGTTIVVDIIVMIGYATLQRIALMIKPKOMKALNKIFGSLFML 194
Db 110 FLDLSKQPWIOYLIWATWTIDLIWAGYTGASKVLRLLRSKQKQKYLNRGFVWFSC 169
QY 195 VGALLASAEHA 205

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:12:37 ; Search time 20 Seconds
(without alignments)
985.963 Million cell updates/sec

Title: US-09-847-392-2
Perfect score: 1061
Sequence: 1 MLEWFPAYLLTSLTSP.....KIFGSLFMLVGLLASARHA 205

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	898.5	84.7	206	2 B91223	homoserine/homoser
2	818.5	77.1	206	2 AF0917	homoserine/homoser
3	698.5	65.8	206	2 AG0466	probable homoserin
4	686	64.7	138	2 A65187	hypothetical 15.4
5	680	64.1	138	2 H86069	hypothetical prote
6	459.5	43.3	205	2 G82358	conserved hypothet
7	389	36.7	209	2 C82990	hypothetical prote
8	256	24.1	212	2 A2805	homoserine/homoser
9	256	24.1	273	2 F97584	hypothetical prote
10	245	23.1	210	2 AC2982	RhtB family transp
11	245	23.1	278	2 E98301	hypothetical prote
12	235	22.1	216	2 AH3203	RhtB family transp
13	228	21.5	241	2 C75329	conserved hypothet
14	208.5	19.7	203	2 A82693	RhtB family transp
15	208.5	19.7	239	2 H7464	conserved hypothet
16	198	18.7	235	2 AD2904	RhtB family transp
17	198	18.7	235	2 F97679	hypothetical prote
18	195.5	18.4	212	2 F64940	hypothetical prote
19	195	18.4	210	2 F69975	dihydrodipicolinat
20	193	18.2	210	2 G83082	hypothetical prote
21	192.5	18.1	207	2 E83703	hypothetical prote
22	191.5	18.0	212	2 C90942	hypothetical prote
23	190	17.9	212	2 G82200	conserved hypothet
24	188	17.7	222	2 E87264	efflux protein, Ly
25	187.5	17.7	212	2 AD0714	probable membrane
26	186.5	17.6	212	2 G85790	hypothetical prote
27	186	17.5	208	2 C82471	conserved hypothet
28	182.5	17.2	214	2 AG2684	RhtB family transp
29	182.5	17.2	224	2 E97466	hypothetical prote

30 180.5 17.0 207 2 D83187
31 180.5 17.0 216 2 F83051
32 178.5 16.8 213 2 C85233
33 178.5 16.8 216 2 C55580
34 172 16.2 249 2 AG3485
35 171 16.1 218 2 A96009
36 168.5 15.9 210 2 E87252
37 167 15.7 212 2 H87498
38 166.5 15.7 205 2 AH3641
39 166.5 15.7 208 2 G87305
40 166 15.6 206 2 C82139
41 163.5 15.4 205 2 AI2614
42 163.5 15.4 224 2 H97396
43 161.5 15.2 210 2 AH2720
44 161.5 15.2 210 2 D97502
45 156 14.7 204 2 B83279

ALIGNMENTS

RESULT 1

B91223
homoserine/homoserine lactone efflux protein [imported] - Escherichia coli (strain O15
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B91223
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: B91223
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA838177.1; PID:gi3364230; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: ECs4754

Query Match 84.7%; Score 898.5; DB 2; Length 206;
Best Local Similarity 89.9%; Pred No. 7.3e-78;
Matches 187; Conservative 3; Mismatches 13; Indels 5; Gaps 3;
Qy 1 MLEWFPAYLLTSLTSPGAINMTTSLNHGYPAGGVYCWASDRG---DSYCAGW 57
Db 1 MLEWFPAYLLTSLTSPGAINMTTSLNHGYPAGGVYCWASDRG---DSYCAGW 58
Qy 58 RGVCTLSRSRVIAFEVLKWAAGAYLWLGICQWRAGADLKSASTQSRHLPORAVFV 117
Db 59 VGLGTLFSRSRVIAFEVLKWAAGAYLWLGICQWRAGADLKSASTQSRHLPORAVFV 118
Qy 118 NLTPKSIIVFLAALFPQFIMFPQQLMQYIVLGVTTIVVDIIIVMIGYATLAQRHIALWIKG 177
Db 119 NLTPKSIIVFLAALFPQFIMFPQQLMQYIVLGVTTIVVDIIIVMIGYATLAQRHIALWIKG 178
Qy 178 PKQKALNKIFGSLFMLVGLLASARHA 205
Db 179 PKQKALNKIFGSLFMLVGLLASARHA 206

RESULT 2

AF0917
homoserine/homoserine lactone efflux protein [imported] - Salmonella enterica subsp. en
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AF0917
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
, S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AF0917
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07932.1; PID:g16504477; GSPDB:GN00176
A:Genetics: rhtB

Query Match 77.1%; Score 818.5; DB 2; Length 206;
Best Local Similarity 78.7%; Pred. No. 2.8e-70;
Matches 166; Conservative 13; Mismatches 21; Indels 11; Gaps 3;

QY 1 MTEWFWAYLLTSILTLSPGSGAINTWTSLNHGYAGGYVCHASDRTGDSYCAGR-- 58
Db 1 MPEWFWAYLLTSTLLSPGSGAINTWTSLNHGY-RGAASAGQGTG----LGIHIV 55
QY 59 ----GVGTLFSSVIAFEVLKWAAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORA 114
Db 56 LVGVGLGTLFSSVIAFEVLKWAAYLIWLGIQOWRAAGAILDLTLAQTSRGLFKRA 115
QY 115 VFVNLNPKSIVFLAALPQFIMPQPOLMQYIVLGVTTIIVDIIWIGVATLAQRIALW 174
Db 116 IFVNLNPKSIVFLAALPQFIMPQPOLMQYIVLGVTTIIVDIIWIGVATLAQRIAAW 175
QY 175 IKGPKQKALNKIFGSLFVLVGLLASARHA 205
Db 176 IKGPKQKALNKAFGSLFVLVGLLASARHA 206

RESULT 3
AG0466
probable homoserine/homoserine lactone efflux protein rhtB [imported] - *Yersinia pestis*
C:Species: *Yersinia pestis*
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AG0466

R:Farhnik, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11596360

A:Accession: AG0466
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC93299.1; PID:g15981746; GSPDB:GN00175
A:Genetics: rhtB

Query Match 65.8%; Score 698.5; DB 2; Length 206;
Best Local Similarity 63.3%; Pred. No. 6.9e-59;
Matches 136; Conservative 33; Mismatches 23; Indels 23; Gaps 3;

QY 1 MTEWFWAYLLTSILTLSPGSGAINTWTSLNHGYAGGYVCHASDRTGDSYCAGR-- 58
Db 1 MTEWFWAYLLTSILTLSPGSGAINTWTSLNHGYAGGYVCHASDRTGDSYCAGR-- 58
QY 59 ----GVGTLFSSVIAFEVLKWAAYLIWLGIQOWRAAGAILDKSLASTQSR 108
Db 50 LAVHIVLVGVGLGALVSQSLAFELTKWGAAYLIWLGIQOWRAAGSLDLHALANSMPFR 109
QY 109 HLFQRAVFNLTNPKSIVFLAALPQFIMPQPOLMQYIVLGVTTIIVDIIWIGVATLA 168
Db 110 KLFKRAVFNLTNPKSIVFLAALPQFIMPQPOLMQYIVLGVTTIIVDIIWIGVATLA 169
QY 169 ORIALWIKPKQKALNKIFGSLFVLVGLLASAR 203
Db 170 TRIARWIKSPQKALNKIFGSLFVLVGLLASAR 204

RESULT 4

A65187
hypothetical 15.4 kD protein in recQ-pldB intergenic region - *Escherichia coli* (strain
N:Alternate names: hypothetical protein f138
C:Species: *Escherichia coli*
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A65187; S30714
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: A65187
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-138 <BLAT>
A:Cross-references: GB:AB000458; GB:U00096; NID:G367299; PIDN:AAC76827.1; PID:g1790258
A:Experimental source: strain K-12, substrain MG1655
R:Daniel, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992

A:Title: Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5
A:Reference number: S30660; MUID:92358234; PMID:1379743
A:Accession: S30714

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'V', 2-138 <DAN>

A:Cross-references: EMBL:M87049
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1992
C:Genetics:
A:Gene: yigK

Query Match 64.7%; Score 686; DB 2; Length 138;
Best Local Similarity 99.3%; Pred. No. 6.8e-58;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VTAEVLKWAAGAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFNLTNPKSIVF 127
Db 1 MTAEVLKWAAGAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFNLTNPKSIVF 60

QY 128 LAALFPQFIMPQPOLMQYIVLGVTTIIVDIIWIGVATLAQRIALWIKGPKQKALNKI 187
Db 61 LAALFPQFIMPQPOLMQYIVLGVTTIIVDIIWIGVATLAQRIALWIKGPKQKALNKI 120

QY 188 FGSFLFVLVGLLASARHA 205
Db 121 FGSFLFVLVGLLASARHA 138

RESULT 5

H86069
hypothetical protein yigK [imported] - *Escherichia coli* (strain O157:H7, substrain EDL9
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H86069

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: H86069
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <STO>

A:Cross-references: GB:AB005174; NID:g13518698; PIDN:AAG59020.1; GSPDB:GN00145; UWGP:25
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yigK

Query Match 64.1%; Score 680; DB 2; Length 138;
Best Local Similarity 99.6%; Pred. No. 2.5e-57;
Matches 136; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 68 VTAEVLKWAAGAYLIWLGIQOWRAAGAILDKSLASTQSRRLHFORAVFNLTNPKSIVF 127

DB 1 MIAFEVLKWAAGAYLWLGIQWRAAGAILDKSLASTQSRRLHFORAVFVNUTNPKSIYF 60
QY 128 LAALFPQFIMPOQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 187
DB 61 LAALFPQFIMPOQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKPKQMKALNKI 120
QY 188 FGSLFMLVGALLASARHA 205
DB 121 FGSLFMLVGALLASAXHA 138
RESULT 6
G82358 conserved hypothetical protein VC0136 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: G82358
R:Reidberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; B
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Basu, S.; Qin, H.; Dragoi, I.; Sellers, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: G82358
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <HEI>
A:Cross-references: GB:AE004104; GB:AE003852; NID:G9654534; PIDN:AAF93313.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0136
A:Map position: 1
C:Superfamily: hypothetical protein b1798
Query Match 43.3%; Score 459.5; DB 2; Length 205;
Best Local Similarity 43.4%; Pred. No. 3.2e-36;
Matches 89; Conservative 44; Mismatches 69; Indels 3; Gaps 2;
QY 1 MTLLEWFAVLLTSLILFLSPGSGAINMTTSLNHG--YPAGGVYCVASDRTGSDSYCAGWR 58
DB 1 MDHVWLAYLLTAVPSLAFSGTVNSINGLSYGRHSLGAILGIGLACHIVLVG-I 59
QY 59 GVGTLFSPRSVIAFEVLKWAAGAYLWLGIQWRAAGAILDKSLASTQSRRLHFORAVFN 118
DB 60 GIGALVAQSALATFLIKWIGAAVLMVIGIQKDRAPLTATTTSHLSQAALLRKAVLIN 119
QY 119 LTPKSIIVFLAALFPQIMQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALWIKGP 178
DB 120 LTPKSIIVFLAALFPQIDTPHWPQFLVIGTTITVDAIVMFGYATLAQGLRVIRSP 179
QY 179 KQKALKNKIFGSLFMLVGALLASAR 203
DB 180 NIMTRMKNLFGSMFMGCGMLLATAK 204
RESULT 7
C82990 hypothetical protein PA5249 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Feb-2002
C:Accession: C82990
R:Stover, C.K.; Yam, X.Q.; Erwin, A.L.; Mitsuuchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Cuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: C82990
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <STO>
A:Cross-references: GB:AE004937; GB:AE004091; NID:G9951553; PIDN:AA08634.1; GSPDB:GN002

F97584
 Hypothetical protein AGR_C3421 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
 C:Accession: F97584
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: F97584
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-273 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK87631.1; PID:gl5156981; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C3421
 A:Map position: circular chromosome

Query Match 24.1%; Score 256; DB 2; Length 273;
 Best Local Similarity 34.1%; Pred. No. 8.8e-17;
 Matches 71; Conservative 29; Mismatches 88; Indels 20; Gaps 6;

QY 1 MTEWFAVLLTIIITLSPGSGAINMTTSLNHGYPAG-----GVYCWASDRDGD--SY 53
 DB 62 MPLENMLFAVAAAIMLAIPGPTILIVISYALGHGKASTATVTGV-----ALGDFTAM 115
 QY 54 CACMRGVTLPFSRSVIAFEVLKWAAGAYLIWLGIOQWRA---AGAILKSLASTQSRRH 109
 DB 116 TSMGLGALLATSAALFTGLKMGAGYLIYIGIKLRSPVGGGA-DGQVGTGREPLK 174
 QY 110 LFORAVFVNLTNPKSIVFLAALFPQFIMPQPOLMQYIVLGVTTIVVDIIVMIGYATLQ 169
 DB 175 IFLHAVIVTALNPKSIVFVFAFLPQFVPLPFPQVLFEFTFVLATVNAALYGLIAS 234
 QY 170 RIALWIKGPKQMKALNKIFGSLFMLVGA 197
 DB 235 AARTNKPKVQRIVNRVTGGGL--LIGA 260

RESULT 10
 AC2982
 RhtB family transporter rhtB [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AC2982
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
 ; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AC2982
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-210 <KUR>
 A:Cross-references: GB:AE008689; PIDN:AAI44273.1; PID:gl7741859; GSPDB:GN00187
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: rhtB
 A:Map position: linear chromosome

Query Match 23.1%; Score 245; DB 2; Length 210;
 Best Local Similarity 38.8%; Pred. No. 7.3e-16;
 Matches 62; Conservative 22; Mismatches 52; Indels 24; Gaps 5;

QY 59 GVGTLFSSRSVIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQ-----SRHLFQR 113
 DB 60 GLGALLAAASEFVGVKVGVCYLVFLGIMLLSRGM--LGALGSTENGAAARSIFLK 117
 QY 114 AVFNLTNPKSIVFLAALFPQFIMPQPOLMQYIVLGVTTIVVDIIVMIGYATL-AQRI- 171

Db 118 SFLVAVTNPKGYLFFSFLPQFIDPAPQIQQAVLALVFAGIDFVFMFGYALLGAQAVR 177
 QY 172 -----ALWIKGPKQMKALNKIFGSLFMLVGLLGAARHA 205
 Db 178 VLKRSKGLW-----LDRICGGALLTLAASLALYRRA 208

RESULT 11
 E98301
 Hypothetical protein AGR_L2738 [imported] - Agrobacterium tumefaciens (strain C58, Cer
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C:Accession: E98301
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tu
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: E98301
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-278 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK89935.1; PID:gl5159890; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L2738
 A:Map position: linear chromosome

Query Match 23.1%; Score 245; DB 2; Length 278;
 Best Local Similarity 38.8%; Pred. No. 9.9e-16;
 Matches 62; Conservative 22; Mismatches 52; Indels 24; Gaps 5;

QY 59 GVGTLFSSRSVIAFEVLKWAAGAYLIWLGIOQWRAAGAILDKSLASTQ-----SRHLFQR 113
 Db 128 GLGALLAAASEFVGVKVGVCYLVFLGIMLLSRGM--LGALGSTENGAAARSIFLK 185
 QY 114 AVFNLTNPKSIVFLAALFPQFIMPQPOLMQYIVLGVTTIVVDIIVMIGYATL-AQRI- 171
 Db 186 SFLVAVTNPKGYLFFSFLPQFIDPAPQIQQAVLALVFAGIDFVFMFGYALLGAQAVR 245

QY 172 -----ALWIKGPKQMKALNKIFGSLFMLVGLLGAARHA 205
 Db 246 VLKRSKGLW-----LDRICGGALLTLAASLALYRRA 276

RESULT 12
 AH3203
 RhtB family transporter rhtB [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C:Accession: AH3203
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,
 ; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AH3203
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-216 <KUR>
 A:Cross-references: GB:AE008687; PIDN:AAI46046.1; PID:gl7743805; GSPDB:GN00188
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: rhtB
 A:Map position: linear chromosome

Query Match 22.1%; Score 235; DB 2; Length 216;
 Best Local Similarity 31.2%; Pred. No. 6.7e-15;
 Matches 63; Conservative 38; Mismatches 89; Indels 12; Gaps 5;

Db 213 PAVLAWLERSFAAEFVALGAKLA 235

Search completed: March 24, 2004, 21:16:42
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:08:46 ; Search time 18 Seconds
(without alignments)
593.021 Million cell updates/sec

Title: US-09-847-392-2

Perfect score: 1061

Sequence: 1 MTLEWFWFAYLLTSILTLSP.....KIFGSLFVLGALLASARHA 205

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	898.5	84.7	206	1	RHTB_ECOLI
2	819.5	77.2	206	1	RHTB_SALTY
3	818.5	77.1	206	1	RHTB_SALTY
4	195.5	18.4	212	1	YRHP_ECOLI
5	195	18.4	210	1	YRHP_BACSU
6	180.5	17.0	216	1	YBF7_PSPA
7	147	13.9	206	1	YGF27_SYNY3
8	134	12.6	206	1	RHTC_SALTY
9	134	12.6	206	1	RHTC_SALTY
10	126.5	11.9	223	1	YAHN_ECOLI
11	126	11.9	195	1	YFIK_ECOLI
12	122	11.5	206	1	RHTC_ECOLI
13	116	10.9	203	1	CHPE_PSPA
14	113	10.7	211	1	YVGA_ECOLI
15	111.5	10.5	206	1	YVGA_ECOLI
16	94	8.9	457	1	PROY_ECOLI
17	92	8.7	228	1	LYSE_COREF
18	92	8.7	456	1	PROY_SALTY
19	89.5	8.4	233	1	LYSE_COREGL
20	87	8.2	210	1	YD07_HAEN
21	81	7.6	732	1	ATFN_ECOLI
22	79.5	7.5	384	1	CYNX_ECOLI
23	79.5	7.5	581	1	FUR4_SCHPO
24	79	7.4	415	1	NQRE_VISCH
25	79	7.4	615	1	NDO_CABEL
26	78.5	7.4	366	1	SPSE_BACSU
27	78.5	7.4	471	1	PPOX_MYXXA
28	78	7.4	1694	1	SN_MOUSE
29	77.5	7.3	225	1	YVGA_AERYH
30	77	7.3	348	1	OPSD_SARXA
31	77	7.3	543	1	SLGT_VIIPA
32	76	7.2	195	1	Y688_RHILO
33	76	7.2	284	1	MODD_ECO57

ALIGNMENTS

RESULT 1

RHTB_ECOLI
ID_RHTB_ECOLI STANDARD; PRT; 206 AA.
AC P27847;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homoserine/homoserine lactone efflux protein.
GN RHTB OR B3824 OR C4746 OR Z5345 OR EC94754 OR SF3902 OR 83853.
OS Escherichia coli.
OS Escherichia coli O6.
OS Escherichia coli O157:H7, and
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334, 623;
[1] SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=52356234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes."
RL Science 257:771-778(1992).
[2] SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P., Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D., Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T., Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
[3] SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=1206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
[4] SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurakawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Tanaka M., Tohe T., Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T., Kuhara S., Shibata T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

Q8ka37 buchnera ap
Q01625 bacillus su
Q64252 homo sapien
P77429 escherichia
P53387 kluyveromyc
P05715 thermoprote
Q10858 mycobacteri
P45064 haemophilus
P11631 oncorhynchu
Q91049 gallus gall
P28873 candida alb
O15040 homo sapien

34 76 7.2 379 1 FLIP_BUCAP
35 75.5 7.1 261 1 OXAI_BACSU
36 75.5 7.1 445 1 IF36_HUMAN
37 75.5 7.1 460 1 ARCD_ECOLI
38 75.5 7.1 566 1 KHT2_KLULA
39 75 7.1 373 1 Y38K_THETE
40 75 7.1 440 1 YJ99_MYCTU
41 74.5 7.0 394 1 FTSW_HAEN
42 74.5 7.0 460 1 NU4M_ONCMY
43 74.5 7.0 504 1 OCLN_CHICK
44 74.5 7.0 564 1 BMRP_CANAL
45 74.5 7.0 1411 1 Y297_HUMAN

RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
 RX MEDLINE=22272406; PubMed=12384590;
 RA Jin O., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
 RA Yu J.;
 RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
 RT through comparison with genomes of Escherichia coli K12 and O157.";
 RL Nucleic Acids Res. 30:4432-4441(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
 RX MEDLINE=22590274; PubMed=12704152;
 RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
 RA Schwartz D.C., Blattner F.R.;
 RT "Complete genome sequence and comparative genomics of Shigella
 RT flexneri serotype 2a strain 2457T.";
 RL Infect. Immun. 71:2775-2786(2003).
 RN [7]
 RP CHARACTERIZATION.
 RC SPECIES=E.coli;
 RX MEDLINE=99313167; PubMed=10386596;
 RA Zakataeva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
 RA Livshits V.A.;
 RT "The novel transmembrane Escherichia coli proteins involved in the
 RT amino acid efflux.";
 RL FEBS Lett. 452:228-232(1999).
 CC -1- FUNCTION: Conducts the efflux of homoserine and homoserine
 CC lactone.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to the rht family.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
 CC frameshift in position 60.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M87049; AAC7620.1; ALT_FRAME.
 DR EMBL; A800458; AAC74827.1; ALT_FRAME.
 DR EMBL; A8016769; AAN83179.1; -.
 DR EMBL; A8005613; AAG59020.1; ALT_INIT.
 DR EMBL; A8005614; -. NOT ANNOTATED_CDS.
 DR EMBL; AF002567; BAB38177.1; -.
 DR EMBL; A8015396; AAN45337.1; -.
 DR EMBL; A8016990; AAP18861.1; ALT_INIT.
 DR PIR; B91223; B91223.
 DR EcoGene; EG11469; rhtB.
 DR InterPro; IPR004778; Homoser_Thr_eff.
 DR InterPro; IPR01123; Lyse.
 DR Pfam; PF01810; Lyse; 1.
 DR TIGRFAMs; TIGR00949; 2A76; 1.
 DR Transport; Transmembrane; Complete proteome.
 FT TRANSMEM 5 25 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 SQ SEQUENCE 206 AA; 22427 MW; 1164F17738509C8C CRC64;

Query Match

84.7%; Score 898.5; DB 1; Length 206;

Best Local Similarity 89.9%; Pred. No. 1.4e-76;
 Matches 187; Conservative 3; Mismatches 13; Indels 5; Gaps 3;
 QY 1 MTEWTFAYLLTSTILSPGSGAINTMTTSLNHGYGAGVYCWASDRTG---DSYCAGW 57
 DB 1 MTEWTFAYLLTSTILSPGSGAINTMTTSLNHGY-RGAVASIAIGLQTGLATHIVLG- 58
 QY 58 RGVLTLSRSVIAFEVLKWAGAAVLWLGICQWRAGAIIDKSLASTQSRHIFQRAV 117
 DB 59 VGLGTLFSRSVIAFEVLKWAGAAVLWLGICQWRAGAIIDKSLASTQSRHIFQRAV 118
 QY 118 NLTPKSTIVFLAALFPQFIMPQPOLMQYIVLGVTTTIVVDIIIVMIGVATLAQRIALWIKG 177
 DB 119 NLTPKSTIVFLAALFPQFIMPQPOLMQYIVLGVTTTIVVDIIIVMIGVATLAQRIALWIKG 178
 QY 178 PKMKALNKIFGSLFVGLVALLASARHA 205
 DB 179 PKMKALNKIFGSLFVGLVALLASARHA 206
 RESULT 2
 RHTB_SALTY
 ID RHTB_SALTY STANDARD; PRT; 206 AA.
 AC Q9L6N6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Homoserine/homoserine lactone efflux protein.
 GN RHTB OR STM3960 OR STM1.30.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OC NCBI TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: Conducts the efflux of homoserine and homoserine lactone
 CC (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to the rht family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF233324; AAF33432.1; -.
 DR EMBL; AE008884; AAU22804.1; -.
 DR StyGene; SG7777; rhtB.
 DR InterPro; IPR004778; Homoser_Thr_eff.
 DR InterPro; IPR001123; Lyse.
 DR Pfam; PF01810; Lyse; 1.
 DR TIGRFAMs; TIGR00949; 2A76; 1.
 DR Transport; Transmembrane; Complete proteome.
 FT TRANSMEM 5 25 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 SQ SEQUENCE 206 AA; 22322 MW; 5089B357D30750F2 CRC64;

```

Query Match          77.2%; Score 819.5; DB 1; Length 206;
Best Local Similarity 78.2%; Pred. No. 3.2e-69;
Matches 165; Conservative 15; Mismatches 20; Indels 11; Gaps 3;

QY 1 MTELEWFAVLLTSLTSPSGAINTMTTSLNHGYPAGVYCWASDRGDSVCAGWR-- 58
DB 1 MTFEWWFAVLLTSLTSLSPSGAINTMTTSLNHGYPAGVYCWASDRGDSVCAGWR 55
QY 59 ----GVGTLFERSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQSRHLFORA 114
DB 56 LVGVGLGTLFERSLIAFEILKWAAGAYLIWLGIQWRAAGAILDLHTLAQTQSRGLFKEA 115
QY 115 VFVNLTPKSIIVFLAALFPQIMPQPOLMOYIVLGVTTIVVDIIVMIGVATLAQRIALW 174
DB 116 IFVNLTPKSIIVFLAALFPQIMPQPOLMOYIVLGVTTIVVDIIVMIGVATLAQRIAAW 175
QY 175 IKGPKQKALNKIFGSLFMLVGLALLASARHA 205
DB 176 IKGPKQKALNKATGSLFMLVGLALLASARHA 206

RESULT 3
RTE SALT
ID RTE SALT STANDARD; PRT; 206 AA.
AC Q823B4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Homoserine/homoserine lactone efflux protein.
GN RHTB OR STY3599 OR T3337.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
RN [2]
RP SEQUENCE FROM N.A.
CC -!- FUNCTION: Conducts the efflux of homoserine and homoserine
CC lactone (by similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the rht family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AL627278; CAD07932.1; -.

```

```

DR EMBL; AE016845; AAC070865.1; -.
DR InterPro; IPR004778; Homoser_Thr_eff.
DR InterPro; IPR001123; LyseE.
DR Pfam; PF01810; LyseE.1.
DR TIGRFAMs; TIGR00949; 2A76; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
SQ SEQUENCE 206 AA; 22308 MW; 40940D3DC0A5CAF2 CRC64;

Query Match          77.1%; Score 818.5; DB 1; Length 206;
Best Local Similarity 78.7%; Pred. No. 4e-69;
Matches 166; Conservative 13; Mismatches 21; Indels 11; Gaps 3;

QY 1 MTELEWFAVLLTSLTSPSGAINTMTTSLNHGYPAGVYCWASDRGDSVCAGWR-- 58
DB 1 MTFEWWFAVLLTSLTSLSPSGAINTMTTSLNHGYPAGVYCWASDRGDSVCAGWR 55
QY 59 ----GVGTLFERSVIAFEVLKWAAGAYLIWLGIQWRAAGAILDKSLASTQSRHLFORA 114
DB 56 LVGVGLGTLFERSLIAFEILKWAAGAYLIWLGIQWRAAGAILDLHTLAQTQSRGLFKEA 115
QY 115 VFVNLTPKSIIVFLAALFPQIMPQPOLMOYIVLGVTTIVVDIIVMIGVATLAQRIALW 174
DB 116 IFVNLTPKSIIVFLAALFPQIMPQPOLMOYIVLGVTTIVVDIIVMIGVATLAQRIAAW 175
QY 175 IKGPKQKALNKIFGSLFMLVGLALLASARHA 205
DB 176 IKGPKQKALNKATGSLFMLVGLALLASARHA 206

RESULT 4
YEAS ECOLI
ID YEAS ECOLI STANDARD; PRT; 212 AA.
AC P76249; O07969; O07971;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yeas.
GN YEAS OR B1798.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1232-1244(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
RA Sivasubramanian S., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map."
RL DNA Res. 3:379-392(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the rht family.

```



```
Query Match 12.6%; Score 134; DB 1; Length 206;
Best Local Similarity 22.1%; Pred. No. 1.9e-05;
Matches 50; Conservative 32; Mismatches 92; Indels 52; Gaps 8;

QY 6 WPAVLLTIIITLSPGCAINTMTTSLNHGYPAG-----GVYCWASDRTGDSYCAWGRV 60
DB 5 PFTVAMVHVALMSGPDFFVFSQTAVERSKEAMGVILGTCV-----MWAGV 55
QY 61 GTLPERSVIAEVLKX-----AGAYLIWLGIOQWRAAGADLKLSTQSRRL-----110
DB 56 ALLGHLII--EKVAWLHTIIMVGGLYLCWGMQMLRGA-----LKKQDAASSPHIELA 109
QY 111 -----FQRAVFNLNPKSIVFLAALPFPQFTMPQPOLMOCVIVLGT-----TIIVD 157
DB 110 QSGRFLKGLLTNLSNPKAIYFSGVSLFVGDNVGAARWGIFALITLTLELAWFTVVAS 169
QY 158 IIVMTGYATLAQRIALMTKQPKOMKALNKIFGSLFVGLVGLALLASAR 203
DB 170 LPALPKMRGQYQRLAKWIDG-----FAGALFAGFGIHLIISR 206

RESULT 10
YAHN_ECOLI
ID YAHN_ECOLI STANDARD; PRT; 223 AA.
AC P75693; P71307;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yahn.
GN YAHN OR B0328.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
RA Lin D., Nanath A., Oefner S., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the rht family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000140; AAC72431.1; -
CC DR EMBL; U73857; AA818053.1; -
CC DR PIR; H64759; H64759.
CC DR EcoGene; EG13598; yahn.
CC DR InterPro; IPR004778; Homoser_Thr_eff.
CC DR InterPro; IPR001123; LysE.
CC DR Pfam; PF01810; LysE; 1
CC DR TIGRfams; TIGR00249; 2A76; 1.
CC KW Hypothetical protein; transmembrane; Complete proteome.
FT TRANSMEM 22 42
FT TRANSMEM 59 79
FT TRANSMEM 79 79

Query Match 11.9%; Score 126.5; DB 1; Length 223;
Best Local Similarity 21.9%; Pred. No. 0.00011;
Matches 44; Conservative 48; Mismatches 94; Indels 15; Gaps 6;

QY 1 MTEWFPAYLIT---SIITLSPGCAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAW 57
DB 12 ITMDPLHAYLTVLGLFVITFEFGANLFFVVQTSLASGRR--GVLTGLGLGVALGDAFYSL 70
QY 58 R--GVGTLFSPRSVIAEVLKXAGAYLIWLGIOQWRAAGADLKLSTQSRRL--LPQR 113
DB 71 GLFLGLALITTCCEIFSLIRIVGGAYLLWFAWCMRRQSTPQMTLQQQFISAPWYFFRR 130
QY 114 AVFVNLNPKSIVFLAALPFPQFTMPQOP---QLMQYIVLGVTTIVVDIIVMIGYATLAQ 170
DB 131 GLITDLNPKQTVLFFSIFSVTLNAETFTWARLMAWAGIVLASIIWRVLSQAFSLPAVR 190
QY 171 IALWIKGPKQMKALNKIFGSL 191
DB 191 RAY----GRMORVASRVIGAI 207

RESULT 11
YFIK_ECOLI
ID YFIK_ECOLI STANDARD; PRT; 195 AA.
AC P38101;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfik.
GN YFIK OR B2578.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Nashimoto H.;
RT "Non-ribosomal proteins affecting the assembly of ribosomes in
RT Escherichia coli."
RL (In) Nierhaus K.H. (eds.);
RL The translational apparatus, pp.185-195, Plenum Press,
RL New York (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Nashimoto H., Saito N.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Sato Y., Sivasundaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
```



```
DR EMBL; M30198; -; NOT ANNOTATED_CDS.
DR PIR; A98223; A98223.
DR EcoGene; EG11468; rhtC.
DR InterPro; IPR004778; Homoser_Thr_eff.
DR Pfam; PF01810; LyseE.
DR TIGRfam; TIGR00949; 2A76; 1.
KW Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 150 173 POTENTIAL.
SQ SEQUENCE 206 AA; 22474 MW; F64017878CC6D50D CRC64;

Query Match 11.5%; Score 122; DB 1; Length 206;
Best Local Similarity 21.0%; Pred. No. 0.00025;
Matches 46; Conservative 36; Mismatches 89; Indels 48; Gaps 8;

QY 11 LTSIILTLSPGSGAINTWTSLNHGYAG-----GVYCWASDRGTGDSYACAGWGVGTLFS 65
DB 10 MVHIVALMSPGDPFFVQTAVSRSRKEAMVMGVLTGCV-----MVMAGIALGL 60
QY 66 RVIAFEVLKW-----AGAAVLIWLGICQWRAAGAIIDKSLASTOSRRL-----FQ 112
DB 61 HLII--ERWMLHTLIMVGGGLYLCWMGYQMLR--GALKKEAVSAPAPQVELAKSGRSFL 116
QY 113 RAVFVNLTPKSIIVFLAALFPQFIMPPQPOLMQVILGVT-----TIVVDIIVMIGY 164
DB 117 KGLLTNLNLPRAIIVFGVSFLFVGDVNGVTARGIIFALIIVETLAWFTVVASLPAIPQM 176
QY 165 ATLQRIALWIKPKQMKALNKIFGSLFMLVGLALLASAR 203
DB 177 RRGYQRLAKWDG-----FAGALFAGFGIHLIISR 206

RESULT 13
CHPE_PSEAE STANDARD; PRT; 203 AA.
AC O87005;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chemotactic transduction protein chpE.
GN CHPE OR PA0417.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RA Whitchurch C.B., Young M.D., Hobbs M., Mattick J.S.;
RT "Pseudomonas aeruginosa chemotactic transduction genes pill, chpA
RT chpB and downstream genes chpC, chpD and chpE.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.I., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the rht family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
-----
CC EMBL; U79580; AAC23935.1; -.
CC EMBL; AE004479; AAG03806.1; -.
DR PIR; T30317; T30317.
DR InterPro; IPR001123; LyseE.
DR Pfam; PF01810; LyseE; 1.
KW Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 46 66 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
SQ SEQUENCE 203 AA; 21290 MW; 195553C048AAD099 CRC64;

Query Match 10.9%; Score 116; DB 1; Length 203;
Best Local Similarity 30.2%; Pred. No. 0.0009;
Matches 39; Conservative 21; Mismatches 63; Indels 6; Gaps 4;

QY 6 WFAYLLTSIILTLSPGSGAINTWTSLNHGY-PAGGVYCWASDRGTGDSYCA--GWFGVGT 62
DB 5 FLAALLFGFAFNVSPGAVFSETLRRLGTGGRFA--LLVQLGSLIGDAYWALLGLTGLAL 62
QY 63 LFSRSVTAFEVLKWAGAAVLIWLGICQWRAAGAIIDKSLASTOSRRLFORAVFVNLTP 122
DB 63 LLGVEQVRIP-LTLCAAYLANLQVQLRDANSFPPLAEDAGEGGRNATGAGAAISLSNP 121
QY 123 KSIYFLAAL 131
DB 122 KNVYVWGAL 130

RESULT 14
YGGG_ECOLI STANDARD; PRT; 211 AA.
AC P11667;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yggA.
GN YGGG OR B2923.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[2]
RN SEQUENCE OF 15-211 FROM N.A.
RC STRAIN=K12 / CS520;
RX MEDLINE=89313302; PubMed=2546007;
RA Alefounder P.R., Baldwin S.A., Perham S.A., Short N.J.;
RT "Identification, molecular cloning and sequence analysis of a gene
RT cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-
RT phosphoglycerate kinase and a putative second glyceraldehyde 3-
RT phosphate dehydrogenase of Escherichia coli.";
RL Mol. Microbiol. 3:723-732(1989).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U28377; AAA69030.1; -;
DR EMBL; A5000375; AAC75960.1; -;
DR EMBL; X14436; CAA32607.1; -;
DR PIR; B65077; QOEC5A.
DR EcoGene; EG11159; YGGA.
DR InterPro; IPR004777; Lys exporter.
DR Pfam; PF01810; LysE.
DR TIGRFAMs; TIGR00948; 2a75; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 179 199 POTENTIAL.
SQ SEQUENCE 211 AA; 23175 MW; 2DAFE27B6A9BE822 CRC64;

Query Match 10.7%; Score 113; DB 1; Length 211;
Best Local Similarity 24.9%; Pred. No. 0.0018;
Matches 52; Conservative 36; Mismatches 95; Indels 26; Gaps 10;

QY 7 FAX-----LLTSILTLSPGSGAINTWTSINHCYPAG-GVYCWASDRGDSVCAGWRG 59
DB 2 FSYFQGLGALGAMILFLGPQNAFV--MNQGRQYHIMIALCAISDLV--LICAGIFG 57

QY 60 VGTIFRSVIAFEVLKWAAGAYLIWLGIQWRAGAILDKSLAST---QSRRLFORAV 115
DB 58 GSALLMQSPWLLALVTGCVAVLLYVGFAPKTAWSNIE-LASAEVWKQGRWKIATML 116

QY 116 FVNLTNP-----KSIVFLAALFPQFTMPQQLMQYIVLGVTTIVVDIIVMIGVATLAQRI 171
DB 117 AVTNLNPVHYLDTFVVLGSLGQ--LDVEPK--RWFALG--TISAFWFFGLALLAAIL 170

QY 172 ALWIKGPKQKALNKIFGSLFMLVGALLA 200
DB 171 APLRLTAKAQRILNLVGVGVWVFIALQLA 199

RESULT 15
YGGA_AERSA STANDARD; PRT; 206 AA.
AC P70775;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yggA.
GN YGGA.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales;
OC Aeromonadaceae; Aeromonas.
OX NCBI_TaxID=645;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 1102;
RX MEDLINE=97431471; PubMed=9286976;
RA Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,
RT Chhabra S.R., Williams P., Macintyre S., Stewart G.S.A.B.,
RT Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida:
RT identification of the LuxRI homologs AhvRI and AsvRI and their
RT cognate N-acylhomoserine lactone signal molecules.";
RL J. Bacteriol. 179:5271-5281(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U65741; AAB70019.1; ALT INIT.
DR InterPro; IPR004777; Lys exporter.
DR PIR; B65077; QOEC5A.
DR EcoGene; EG11159; YGGA.
DR InterPro; IPR004777; Lys exporter.
DR Pfam; PF01810; LysE.
DR TIGRFAMs; TIGR00948; 2a75; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
SQ SEQUENCE 206 AA; 21505 MW; D1C2C492CDA0179A CRC64;

Query Match 10.5%; Score 111.5; DB 1; Length 206;
Best Local Similarity 24.7%; Pred. No. 0.0024;
Matches 40; Conservative 26; Mismatches 65; Indels 31; Gaps 6;

QY 59 GVGTFERSVIAFEVLKWAAGAYLIWLGIQ---WRAAGAILDKSLASTQSRRLFORA 114
DB 57 GGANLLAASPGLALLTWGGVFLICWFGIRSLRSAMQCGA-----ALADSPRLMGVKS 110

QY 115 VF-----VNLTNP-----KSIVFLAALFPQFTMPQQLMQYIVLGVTTIVVDIIVMIGVA 165
DB 111 VLAHTLGVTLNPHVYLDLMLLGSFGSQFAELRP-----AFAAGAMLASLVNWFYSLA 164

QY 166 TLAQRIALWIKGPKQKALNKIFG-----SLFMLVGALLAS 201
DB 165 FGAAALSPWLAARGVRVQAIDTIVGLMLGLALQALASGALLAS 206

Search completed: March 24, 2004, 21:15:12
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 24, 2004, 21:12:07 ; Search time 45 Seconds
(without alignments)
1437.361 Million cell updates/sec

Title: US-09-847-392-2
Perfect score: 1061
Sequence: 1 MTELEWFAYLITSLTSP.....KIFGSLFVLVGALLASARHA 205

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:**

1: sp_archaea:**

2: sp_bacteria:**

3: sp_fungi:**

4: sp_human:**

5: sp_invertebrate:**

6: sp_mammal:**

7: sp_mhc:**

8: sp_organelle:**

9: sp_phage:**

10: sp_plant:**

11: sp_rodent:**

12: sp_virus:**

13: sp_invertebrate:**

14: sp_unclassified:**

15: sp_virus:**

16: sp_bacteriap:**

17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	698.5	65.8	206	16	Q8ZAH0	Q8ZAH0 yersinia pe
2	470.5	44.3	207	16	Q8DD72	Q8dd72 vibrio vuln
3	459.5	43.3	205	16	Q9KVK7	Q9kvk7 vibrio chol
4	457.5	43.1	205	16	Q87KK4	Q87kk4 vibrio para
5	389	36.7	209	16	Q9HTU5	Q9htu5 pseudomonas
6	367.5	34.6	210	16	Q8B79	Q8b79 pseudomonas
7	354.5	33.4	210	16	Q8B79	Q8b79 pseudomonas
8	312.5	29.5	209	16	Q7W0G4	Q7w0g4 bordetella
9	311.5	29.4	209	16	Q7W0S2	Q7w0s2 bordetella
10	311.5	29.4	209	16	Q7W2S4	Q7w2s4 bordetella
11	294	27.7	203	16	Q8B700	Q8b700 pseudomonas
12	285.5	27.0	211	16	Q87PX1	Q87px1 vibrio para
13	281	26.5	211	16	Q92NK8	Q92nk8 rhizobium m
14	269	25.4	207	16	Q8EKH1	Q8ekh1 shewanella
15	258	24.3	208	2	Q8KW80	Q8kw80 ruegeria sp
16	256	24.1	273	16	Q8UE96	Q8ue96 agrobacteri

17	255.5	24.1	204	16	Q98GT1	Q98gt1 rhizobium 1
18	252.5	23.8	209	16	Q7WQR5	Q7wqr5 bordetella
19	252.5	23.8	209	16	Q7W1T8	Q7wt18 bordetella
20	248	23.4	202	16	Q7W4X4	Q7w4x4 bordetella
21	248	23.4	207	16	Q8KC20	Q8kc20 chlorobium
22	247	23.3	211	16	Q92LY9	Q92ly9 rhizobium m
23	245	23.1	278	16	Q8UAB4	Q8uab4 agrobacteri
24	244.5	23.0	209	16	Q7VS75	Q7vs75 bordetella
25	241.5	22.8	218	16	Q7WGF2	Q7wgf2 bordetella
26	235.5	22.2	208	16	Q983T1	Q983t1 rhizobium 1
27	235	22.1	209	16	Q8B7R5	Q8b7r5 pseudomonas
28	235	22.1	234	16	Q8UJW7	Q8ujw7 agrobacteri
29	228.5	21.5	207	16	Q98CS3	Q98cs3 rhizobium 1
30	228	21.5	241	16	Q9RSX2	Q9rsx2 deinococcus
31	222	20.9	213	16	Q8B7U2	Q8b7u2 pseudomonas
32	221.5	20.9	208	16	Q8B7U2	Q8b7u2 pseudomonas
33	218.5	20.6	203	16	Q98MB1	Q98mb1 rhizobium 1
34	214.5	20.2	203	16	Q7WJY4	Q7wjy4 bordetella
35	214.5	20.2	203	16	Q7WAS5	Q7was5 bordetella
36	212.5	20.0	203	16	Q7VVP8	Q7vvp8 bordetella
37	212	20.0	220	16	Q8XZ84	Q8xz84 ralstonia s
38	211.5	19.9	206	16	Q8B1H4	Q8b1h4 pseudomonas
39	210.5	19.8	211	17	Q8SPF6	Q8spf6 methanosarc
40	210	19.8	210	16	Q81QR4	Q81qr4 bacillus an
41	209.5	19.7	206	16	Q8Y2B8	Q8y2b8 ralstonia s
42	208.5	19.7	239	16	Q8UH19	Q8uh19 agrobacteri
43	208	19.6	210	16	Q81DR4	Q81dr4 bacillus ce
44	206	19.4	208	16	Q8BKW8	Q8bkw8 pseudomonas
45	205.5	19.4	203	16	Q89DU7	Q89du7 bradyrhizob

ALIGNMENTS

RESULT 1

Q8ZAH0	Q8ZAH0	PRELIMINARY;	PRT;	206 AA.
AC	Q8ZAH0;			
DT	01-MAR-2002 (TREMREL. 20, Created)			
DT	01-MAR-2002 (TREMREL. 20, Last sequence update)			
DT	01-JUN-2003 (TREMREL. 24, Last annotation update)			
DE	Putative homoserine/homoserine lactone efflux protein.			
GN	RHTB OR YPO3831 OR Y0399.			
OS	Yersinia pestis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Yersinia.			
OX	NCBI_TaxID=632;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CO-92 / Biovar Orientalis;			
RX	MEDLINE=21470413; PubMed=11586360;			
RA	Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,			
RA	Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,			
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,			
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,			
RA	Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,			
RA	Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,			
RA	Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;			
RT	"Genome sequence of Yersinia pestis, the causative agent of plague."			
RL	Nature 413:523-527(2001).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KIM5 / Biovar Mediaevalis;			
RX	MEDLINE=22137863; PubMed=12142430;			
RA	Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,			
RA	Ferna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,			
RA	Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,			
RA	Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,			
RA	Perry R.D.;			
RT	"Genome sequence of Yersinia pestis KIM."			
RL	J. Bacteriol. 184:4601-4611(2002).			
DR	EMBL; AJ414159; CAC93299.1; -			
DR	EMBL; AE013640; AAM83988.1; -			

```
DR PIR; AG0466; AG0466.
DR GO; GO:0018020; Cmembrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR InterPro; IPR001123; LyseE.
DR Pfam; PF01810; LyseE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 206 AA; 22321 MW; 51C3B7B259992435 CRC64;

Query Match 65.8%; Score 699.5; DB 16; Length 206;
Best Local Similarity 63.3%; Pred. No. 9.8e-58;
Matches 136; Conservative 33; Mismatches 23; Indels 23; Gaps 3;

QY 1 MLEWTFAYLLTSIIILTLSPSGAINTWTSINLHGPAGGYCWASDRTGDSYCAWR-- 58
DB 1 MLDWVLLVTLTLISLSPSGAINTWTSISG--TRGVV-----ASIGGLQLG 49

QY 59 -----GVGTLFSRSVIAFEVLKWAAGAYLIWLGQQWRAAGAILDKSLASTQSR 108
DB 50 LAVHIVLGVGLGALVQSLLAFETLKLGAAYLIWLGQQWRAAGSLDLHALANSPPR 109

QY 109 HLPORAVFVNLNPKSIVFLAALFPQTFMPOQPOLMOYIVLVGTTIVVDIIVMIGVATLA 168
DB 110 KLFKRAVFNLNPKSIVFLAALFPQTFMPOQPOLMOYIVLVGTTIVVDIIVMIGVATLA 169

QY 169 QRALWIKPKQKQKALNKFGLSLFVLGALLASAR 203
DB 170 TRIARWIKSPQKMLNKFGLSLFVLGALLATAR 204

RESULT 2
Q8DD72 PRELIMINARY; PRT; 207 AA.
AC Q8DD72;
DT 01-MAR-2003 (TremBrel. 23, Created)
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE Putative threonine efflux protein.
GN Wll138.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.B.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016800; AAO09612.1;
DR GO; GO:0018020; Cmembrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR InterPro; IPR001123; LyseE.
DR Pfam; PF01810; LyseE; 1.
KW Complete proteome.
SQ SEQUENCE 207 AA; 21859 MW; ADA3B5E7F57E14E2 CRC64;

Query Match 44.3%; Score 470.5; DB 16; Length 207;
Best Local Similarity 44.9%; Pred. No. 2.7e-36;
Matches 97; Conservative 39; Mismatches 45; Indels 35; Gaps 5;

QY 6 WFAVLLTSIIILTLSPSGAINTWTSINLHGPAGGYCWASDRTGDSYCAWR----- 58
DB 8 WLAIVVTLVFLSLAPSGTIVNSISGLSYG-----TRKSLASIVGLQIGLAVHI 56

QY 59 -----GVGTLFSRSVIAFEVLKWAAGAYLIWLGQQWRAAGAILDKSLASTQSR 109
DB 57 VLVGAGIGALVAQSATATVTKWGAAYLVWLGQKWR-----DTSSLA-TASQQAIS 110

QY 110 --LFORAVFVNLNPKSIVFLAALFPQTFMPOQPOLMOYIVLVGTTIVVDIIVMIGVATL 167
DB 111 TALLRKAVLINLTPKSIIVFLVALFPQTFIDPSQDQVQLAVLGIITTVIDAFVNLGYTTL 170
```

```
DB 168 AQRIALWIKPKQKQKALNKFGLSLFVLGALLASAR 203
DB 171 ASQLGRFIRSEKVKWKINKVFGSMFGCGALLAAK 206

RESULT 3
Q9KVK7 PRELIMINARY; PRT; 205 AA.
AC Q9KVK7;
DT 01-OCT-2000 (TremBrel. 15, Created)
DT 01-OCT-2000 (TremBrel. 15, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE Hypothetical protein VCO136.
GN VCO136.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
DR EMBL; AE004104; AAF93313.1; --
DR PIR; G82358; G82358.
DR TIGR; VCO136; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR InterPro; IPR001123; LyseE.
DR Pfam; PF01810; LyseE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 205 AA; 22099 MW; AE42CAG6A5EA1223 CRC64;

Query Match 43.3%; Score 459.5; DB 16; Length 205;
Best Local Similarity 43.4%; Pred. No. 2.9e-35;
Matches 89; Conservative 44; Mismatches 69; Indels 3; Gaps 2;

QY 1 MLEWTFAYLLTSIIILTLSPSGAINTWTSINLHGPAGGYCWASDRTGDSYCAWR 58
DB 1 MDHIVLWLYLLTAVVFLSLAPSGTIVNSISGLSYGTRHSLGAIIGLQIGLACHIVLVG-I 59

QY 59 GVGTLFSRSVIAFEVLKWAAGAYLIWLGQQWRAAGAILDKSLASTQSRHLPQRAVFN 118
DB 60 GIGALVAQSALAFLLIKWGAAYLVWLGQKWRDRAPLTATTSHLSQAALLRAVLIN 119

QY 119 LTNPKSIVFLAALFPQTFMPOQPOLMOYIVLVGTTIVVDIIVMIGVATLAQRALWIKGP 178
DB 120 LTNPKSIVFLVALFPQTFIDTRDHPQFLVLGTTIVVDIIVMFGYTAALAAQLGIRSP 179

QY 179 KQKALNKFGLSLFVLGALLASAR 203
DB 180 NIMTRMNKLFSGSMFGCGMLLATK 204

RESULT 4
Q87KK4 PRELIMINARY; PRT; 205 AA.
ID Q87KK4
AC Q87KK4;
DT 01-JUN-2003 (TremBrel. 24, Created)
DT 01-JUN-2003 (TremBrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBrel. 25, Last annotation update)
```



```
Db 45 GLQIALVAQIAVAAGLGAVALAASEMAFTLIKWFVGAVLYLGIKQWRASPTDLADESAV 104
QY 102 ASTQSRHLFORAVFVNLNPKSIYFLAALPQFTMPQPOLMQYIVLVGTTIVVDIIVM 161
Db 105 RPVKQWTLVFGFLVNLNPKSIYFLAALPQFTMPQPOLMQYIVLVGTTIVVDIIVM 164
QY 162 IGVATIAQRIALWIKGPKQKALNKIFGSLFVLVGLALLAS 201
Db 165 AGYTGGLASKVLKALTKTPQORRLNRTFASLFVGAAGFLAT 204

RESULT 7
Q88RD4 PRELIMINARY; PRT; 210 AA.
AC Q88RD4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Transporter, Lyse family.
GN P01198.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzaz A.,
RA Unterback T., Rizzo M., Lee K., Kosack D., Moesti D., Wedler H.,
RA Lauber J., Stjepandic D., Rohselz J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Tammis K.N., Dueterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-806(2002).
DR EMBL; AE016774; AAN65831.1; -.
DR TIGR; P01198; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006865; P:amino acid transport; IEA.
DR InterPro; IPR001123; LyseE.
DR Pfam; PF01810; LyseE; 1.
DR Complete proteome.
SQ SEQUENCE 210 AA; 22827 MW; 89B800A87428770A CRC64;

Query Match 33.4%; Score 354.5; DB 16; Length 210;
Best Local Similarity 32.3%; Pred. No. 2.2e-25;
Matches 74; Conservative 41; Mismatches 69; Indels 45; Gaps 4;

QY 1 MTLWWFAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGW-- 58
Db 1 MSNEVWLGFAACWVISLSPGASMSGLQYF-----WRGY 40
QY 59 -----GVGTLFERSVIAFEVLKWAGAAVLIWLGICQWRAAGA-- 99
Db 41 WNALGLQLGLIMQIAIIAAGVAVLAASATAFQIKWFGVLYLVAYKQWRAL-PYDMS 99
QY 100 SLASTQ---SRHLFORAVFVNLNPKSIYFLAALPQFTMPQPOLMQYIVLVGTTIVV 156
Db 100 DSGVRPIGKPLSLVFRGLVNLNPKSIYFLAALPQFTMPQPOLMQYIVLVGTTIVV 159
QY 157 DIIVMIGVATLQRIALWIKGPKQKALNKIFGSLFVLVGLALLASARHA 205
Db 160 DLLVWAGYGLASHVLMRLRTPKQKRLNRTFAGLFGAATFLATLRR 208

RESULT 8
Q7W0G4 PRELIMINARY; PRT; 209 AA.
ID Q7W0G4
```

```
AC Q7W0G4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lyse family efflux protein.
GN B01171.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Gobie A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640411; CAE40550.1; -.
DR Complete proteome.
SQ SEQUENCE 209 AA; 22779 MW; F77DDE3CDD41B54 CRC64;

Query Match 29.5%; Score 312.5; DB 16; Length 209;
Best Local Similarity 33.5%; Pred. No. 2e-21;
Matches 73; Conservative 39; Mismatches 79; Indels 27; Gaps 5;

QY 1 MTLWWFAYLLTSLTSPGSGAINTMTTSLNHGYPAGGYCWASDRTGDSYCAGW-- 57
Db 1 MTLSTWLTFFVSWAISPSFGAGAISAMSSGLKVGARG---YWN-----AGLILG 49
QY 58 -----RGVTLFERSVIAFEVLKWAGAAVLIWLGICQWRAAGA-- 105
Db 50 ILFOFLIVAVGLGAVLATSELAFTLVKYAGVWVLIYLGMRQIRTDAAPTVDAGD-PHRA 108
QY 106 SRRHLFORAVFVNLNPKSIYFLAALPQFTMPQPOLMQYIVLVGTTIVVDIIVMIGVA 165
Db 109 SIRELVGRGLINTMNPQGVFLAVVPQVDFPAQLTQQLALAGTLAFTLDVANDVT 168
QY 166 TIAQRIALWIKGPKQKALNKIFGSLFVLVGLALLASAR 203
Db 169 LLAARVLMRLKKAHHIRWNRVFGSLFILAGVFLATFR 206

RESULT 9
Q7WDS2 PRELIMINARY; PRT; 209 AA.
ID Q7WDS2
AC Q7WDS2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Lyse family efflux protein.
GN B04915.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RB50 / ATCC BAA-589;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
```


RESULT 12

Q87PX1 PRELIMINARY; PRT; 219 AA.
AC Q87PX1, (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Putative homoserine/homoserine lactone efflux protein.
GN VP1379.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
[1]
SEQUENCE FROM N.A.
RP STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Nakano M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae.";
RL Lancet 361:743-749 (2003).
DR EMBL; AP005077; BAC59642.1; .
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0008665; P:amino acid transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000005; HTHARAC.
DR InterPro; IPR001123; Lyse.
DR Pfam; PF01810; Lyse; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
KW Complete proteome.
SQ SEQUENCE 219 AA; 23558 MW; FBSD455165DS7500 CRC64;

Query Match 27.0%; Score 286.5; DB 16; Length 219;
Best Local Similarity 29.7%; Pred. No. 5.8e-19;
Matches 65; Conservative 43; Mismatches 86; Indels 25; Gaps 3;

QY 1 MLEWFWAYLLTSIILTPSGGAINMTTSLNHGYPAGGYCWASDRGDSVCAG-----56
DB 12 MQLDTWYITLAILVLTASPGPSLLCLSKGVSSGF-----RLALTALGSLA 60
QY 57 -----WRGVGTLFSSVIAFEVLKWAAGAYLWLGIOQWRAAGAILDKSLASTQSR 108
DB 61 ITIILTLSTGLGVIASEFVFNKWCGRAYLWLGIOAFRSKQNDFAKSDSAQVSTS 120
QY 109 HL--FORAVFVNLTPKSIIVFLAALFPQFIMPQPOLMQYIVLGVTTIVVDIIVMIGYAT 166
DB 121 HVSAYTSGFVGSNNPKAIIFFTLFPQFIDPTASLLTQVAIFAGTFVTFELSLWTFYAL 180
QY 167 LAQRIALWTKGPKOMKALNKIFGSLFVLVGLLASARHA 205
DB 181 LGVKSINWLFPEAGRAKLFNLTGGVFISAGVNLSTANES 219

RESULT 13

Q92NK8 PRELIMINARY; PRT; 211 AA.
AC Q92NK8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative amino acid efflux protein.
GN R02191 OR SMC01851.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;

[1]

RN SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kias E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ransperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Gallibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).
DR EMBL; AL591789; CAC46770.1; .
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005293; F:lysine permease activity; IEA.
DR GO; GO:0006885; P:amino acid transport; IEA.
DR InterPro; IPR001123; Lyse.
DR Pfam; PF01810; Lyse; 1.
KW Complete proteome.
SQ SEQUENCE 211 AA; 22293 MW; 7955F42044DBCF47 CRC64;

Query Match 26.5%; Score 281; DB 16; Length 211;
Best Local Similarity 33.8%; Pred. No. 1.8e-18;
Matches 71; Conservative 27; Mismatches 106; Indels 6; Gaps 3;

QY 1 MLEWFWAYLLTSIILTPSGGAINMTTSLNHGYPAGGYCWASDRGDSVCAG-----58
DB 1 MSFEHMFAPAAASAVLLAIPGPTILLVISVALGCHGRKIAGATV-AGVALGDFMTAMTASML 59
QY 59 GVGTLFSSVIAFEVLKWAAGAYLWLGIOQWRAAGAILDKSLASTQSR---HLFQRAV 115
DB 60 GLGALLATSAAVFTVLKWIAGAYLWLGKLPAPVNDGSGTSPAEPRIPRIFLHTY 119
QY 116 FVNLTNPKSIVFLAALFPQFIMPQPOLMQYIVLGVTTIVVDIIVMIGYATLAQRIALW 175
DB 120 AVTALNPKSILFFVAFLPQFLDLRLPFAQWAFETFTLILATINALVLAALAAAGSTI 179
QY 176 KGPQMKALNKIPGSLFVLVGLLASARHA 205
DB 180 KPNIRIRVARELGGSLIGAGFLTAGLKRA 209

RESULT 14

Q8EKH1 PRELIMINARY; PRT; 207 AA.
AC Q8EKH1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Transporter, Lyse family.
GN S00122.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=MR-1.
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Mavuthavan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015463; AAN53209.1; .
DR TIGR; S00122; .
DR GO; GO:0016020; C:membrane; IEA.

